AD_____

Award Number: W81XWH-04-1-0588

TITLE: Aurora-A as a Modifier of Breast Cancer Risk in BRCA1/2 Mutation Carriers

PRINCIPAL INVESTIGATOR: Fergus J. Couch, Ph.D.

CONTRACTING ORGANIZATION: Mayo Clinic
Rochester, MN 55905

REPORT DATE: June 2007

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-06-2007	Final	21 May 2004 – 20 May 2007

4. TITLE AND SUBTITLE

Aurora-A as a Modifier of Breast Cancer Risk in BRCA1/2 Mutation Carriers

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0588

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Fergus J. Couch, Ph.D.

E-Mail: couch.fergus@mayo.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Mayo Clinic
Rochester, MN 55905

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The AURORA-A/BTAK/STK15 gene encodes a centrosome-associated kinase that causes centrosome amplification, failure of cytokinesis, and aneuploidy when amplified and/or overexpressed in breast tumors. A number of gene association studies using matching breast cancer cases and controls have shown that the F31I polymorphism in STK15 increases risk of breast. We hypothesized that the F31I polymorphism is associated with increased risk of breast cancer in BRCA1 and BRCA2 mutation carriers.Using over 7,000 carriers of *BRCA1* and *BRCA2* deleterious mutations with a mean age of breast cancer diagnosis of 42 years we have shown that F31I has no influence on breast cancer risk. In parallel, we have completed an association study of 2400 polymorphisms in other cell division regulatory genes in 800 breast cancer cases and 800 controls collected at the Mayo Clinic. A total of 144 polymorphisms displayed significant associations with breast cancer risk. An effort to validate these findings in 4,000 *BRCA1* and *BRCA2* mutation carriers from six collaborating groups has been initiated, but this is not yet complete.

15. SUBJECT TERMS
Breast Cancer, BRCA1, BRCA2, Aurora-A, genetic modifier

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	29	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Introduction

Studies have shown that the age of onset of breast cancers in BRCA2 mutation carriers is significantly later than for BRCA1 mutation carriers (Fodor et al., 1998). However, while the age specific penetrance may differ the cumulative lifetime risk appears to be similar. In addition, there is substantial variation in the age of onset and the site of cancer amongst BRCA1 and BRCA2 mutation carriers, even in the same family (Goldgar et al., 1994). This strongly suggests that genetic and/or environmental modifiers of breast cancer risk in BRCA1/2 mutation carriers exist (Rebbeck 2002). Certainly, some component of this effect is due to differential risks associated with different mutations in the genes (Gayther et al. 1995, 1997). However, there are likely to be multiple low-penetrance genes that also increase the susceptibility to breast cancer. Mutated forms of these genes probably confer only a small to moderate increase in the lifetime breast cancer risk, but because variations in these low penetrance genes are present in a large number of people, the population risk for breast cancer caused by these genes could be substantial (Rebbeck 1999). These observations raise the question of whether genes associated with other functions of BRCA1 and BRCA2 might also be modifiers of breast cancer risk in carriers.

Recent findings have shown that both BRCA1 and BRCA2 are involved in regulation of the G_2 to M phase transition in the cell cycle. In addition, it has been shown that both proteins are localized to the centrosome and regulate centrosome duplication and centrosome function (Hsu et al., 2001; Nakanishi et al., 2007). Mutations in BRCA1 and BRCA2 are correlated with aberrant duplication of the centrosome leading to centrosome amplification, chromosome mis-segregation, and aneuploidy (Xu et al., 1999; Deng 2002; Starita et al., 2004; Wu et al., 2005). Based on these data, we questioned whether other proteins that mediate centrosome function might act as modifiers of breast cancer risk in BRCA1/2 carriers. The AURORA-A/AURKA/BTAK/STK15 gene encodes a centrosome-associated kinase that causes centrosome amplification, failure of cytokinesis, and aneuploidy when amplified and/or overexpressed in breast tumors. STK15 is also known to bind to BRCA1 and BRCA2 (Ouchi et al., 2004; unpublished data). The F31I polymorphism in STK15 was originally identified as a candidate lung tumor risk modifier locus in a mouse model (Ewart-Toland et al., 2003). F31I altered the activity of the Aurora box-1 of the STK15 protein, resulting in disruption of p53 binding and a decreased rate of degradation of STK15. The stabilized STK15 was associated with centrosome amplification and failure of cytokinesis, increased chromosomal instability and aneuploidy, suggesting a direct effect on the F31I variant on promotion of tumor formation (Ewart-Toland et al., 2003). In a study of incident breast cancer cases (n = 941) and age-matched population controls (n=830), Egan et al. (Egan et al., 2004) found that the breast cancer risk for Ile/Ile homozygotes were at increased risk for breast cancer (OR = 1.54; 95% CI: 0.96-2.47), although this finding was not significant. Sun et al. observed that the Ile encoding allele is the common allele in the Chinese population whereas the Phe encoding allele is more common in Caucasian populations (Sun et al., 2004). In addition, an association between Ile/Ile homozygotes and ER negative breast carcinomas (OR = 2.56; 95% CI: 1.24-5.26) was detected. Lo et al. reported a significant association between *AURKA* haplotypes and breast cancer risk (Lo et al., 2005). Ewart-Toland et al. also found an increase in cancer risk for the *Ile/Ile* homozygotes (OR = 1.35, 95% CI: 1.12–1.64; $p = 0.002$) in a meta-analysis of data from four case–control breast cancer populations (Ewart-Toland et al., 2005). Based on these data, we hypothesized that the F31I polymorphism is associated with increased risk of breast cancer in BRCA1 and BRCA2 mutation carriers.

Since then additional studies of STK15 F31I have been completed. Post-menopausal women homozygous for the F31I and I57V alleles of *AURKA* in a case-control study nested within the Nurses' Health Study prospective cohort had an increased risk of invasive breast cancer (OR 1.63, 95% CI 1.08–2.45) (Cox et al., 2006). In contrast, Dai et al. did not observe a significant association with breast cancer risk for Ile/Ile homozygotes (OR = 1.2; 95% CI, 0.9-1.6) in a population based case-control series of Han Chinese (Dai et al., 2004), and Fletcher et al. (Fletcher et al., 2006) found no association between Ile/Ile homozygotes and risk of bilateral breast cancer (OR = 0.63, 95% CI 0.34-1.13).

Body

Aim 1. To validate the association between Val57Ile in STK15 and increased risk of breast cancer in a large cohort of BRCA1/2 mutation carriers.

Task 1. We began the study by genotyping the F31I and V57I polymorphisms in DNAs from 1332 carriers of BRCA1 and BRCA2 deleterious mutations that were provided by four collaborating groups.

Task 2. No association with risk of breast cancer in BRCA1/2 mutation carriers was observed for F31I heterozygotes (OR = 0.95; 95%CI 0.82-1.11) or for V57I heterozygotes (OR = 1.01; 95%CI 0.86-1.18) or homozygotes (OR = 0.71; 95%CI 0.41-1.24). However, homozygosity for the F31I allele was associated with an increased risk of breast cancer (OR = 1.23; 95%CI 0.93-1.63). This association, while insignificant, was consistent even when evaluating *BRCA1* carriers (OR = 1.24; 95%CI 0.90-1.71) or BRCA2 carriers (OR = 1.20; 95%CI 0.68-2.14) alone.

Task 1. In an effort to verify the association of F31I with breast cancer risk in *BRCA1* and *BRCA2* carriers we established a large consortium of investigators from the U.S.A. and Europe. A total of 4935 female *BRCA1*, 2241 female *BRCA2* deleterious mutation carriers and 11 individuals carrying both *BRCA1* and *BRCA2* mutations from 16 participating groups were included in this study. Of these 7187 mutation carriers, 3884 had a diagnosis of breast cancer at the end of follow up and 3303 were censored as unaffected at a mean age of 43.4 years.

Task 2. To avoid overlap between studies we compared carriers by country of origin, year of birth, mutation and reported ages. The frequency of the recessive Ile/Ile encoding genotype in the 16 groups varied between 3% and 8%, which is similar to estimates from other populations. There was no difference in the frequency of the Ile/Ile recessive genotype across genotyping platforms (p=0.33). Similarly, the study sites with the highest Ile/Ile frequencies did not have ethnic mixtures significantly different to the other study sites. The F31I polymorphism did not deviate significantly from Hardy-Weinberg equilibrium (p=0.07) among all 7187 affected and unaffected carriers.

The estimated risk of breast cancer associated with the recessive genotype for F31I in *BRCA1* and *BRCA2* carriers was calculated using a weighted Cox proportional hazards model. While there was a suggestion of a protective effect (HR = 0.91; 95%CI 0.77-1.06) overall, the result was not statistically significant. Similarly, no association with risk was observed for individual participating centers, other than for two centers that contributed small numbers of carriers to the study. A test for heterogeneity across study site was not significant (p=0.06). We also evaluated whether the Ile/Ile genotype was associated with risk of breast cancer in *BRCA1* carreirs alone or *BRCA2* carriers alone. No significant association with risk was detected for either *BRCA1* (HR = 0.90; 95%CI 0.75-1.08) or *BRCA2* carriers (HR = 0.93; 95%CI 0.67-1.29) (Couch et al., 2007). As other studies have reported an association between the recessive Ile/Ile encoding genotype and postmenopausal status in non-carriers (Egan et al., 2003; Cox et al., 2007), we considered the influence of menopausal status of carriers on breast cancer risk. At the end of follow-up, 4201 carriers were pre-menopausal and 2986 were post-menopausal. No significant association with risk was detected (Couch et al., 2007). Because prophylactic oophorectomy substantially reduces the risk of breast cancer in *BRCA1* and *BRCA2* mutation carriers (REbbeck et al., 2002), we also evaluated the influence of prophylactic oophorectomy status. A total of 707 individuals reported undergoing prophylactic oophorectomy, 4298 reported no history of oophorectomy, while 2182 (30%) provided no data at last follow up. Associations with breast cancer risk by category of prophylactic oophorectomy did not differ markedly from the overall results. Secondary analyses using a two degree-of-freedom general model also failed to detect a significant association for either a single copy (p=0.97) or two copies (p=0.24) of the F31I polymorphism compared to no copies (Couch et al., 2007).

In an effort to account for possible survival bias and the inclusion of prevalent cases in the collection of *BRCA1* and *BRCA2* carriers, we repeated our analysis after excluding cases diagnosed more than three years prior to the date of ascertainment. For this analysis we excluded records where an age at interview was not provided. Overall, the mean difference between age of diagnosis and age at interview for the 3422 cases with available

data was 8.7 years. Of these 1,322 (38.6%) cases had been diagnosed less than three years prior to the date of ascertainment. When excluding prevalent cases no association between the Ile/Ile genotype and breast cancer risk was observed, and the risk estimates were similar to those obtained when using both prevalent and incident cases (Couch et al., 2007). Thus, STK15 F31I does not appear to be associated with breast cancer risk modification in *BRCA1* or BRCA2 mutation carriers.

In parallel, we worked with a separate consortium of investigators to assess the influence of STK15 F31I on breast cancer risk in sporadic breast cancer cases. This consortium, named Breast Cancer Association Consortium (BCAC) is comprised of 18 groups from the USA and Europe who are pooling genotyping data on various polymorphisms in order to generate sufficient sample sizes for clarifying genetic risks associated with these polymorphisms. We genotyped the F31I polymorphism on 724 breast cancer cases and 767 controls. Cases were collected through the Mayo Clinic oncology clinic from 2002 to 2005 and were restricted to Caucasians from a 6-state region surrounding the Mayo Clinic. Controls were recruited from Internal Medicine Clinics at Mayo Clinic, had no previous history of breast cancer and were matched to cases by age and residence. In the Mayo Clinic case-control study the F31I polymorphism was not associated with altered risk of breast cancer (OR = 1.00 (0.80-1.24)) for heterozygotes and (OR = 0.95 (0.59-1.52)) for homozygotes. Similarly when pooled with data from five other centers no association with risk was observed (OR = 0.98 (0.92-1.04)) for homozygotes. Stratifying by age in order to consider postmenopausal women only also failed to identify any association with risk. This report completes all effort associated with Tasks 1-2.

Task 3. In Aim #1 we also stated that we would evaluate single nucleotide polymorphisms (SNPs) in other mitotic regulators for effects on breast cancer risk in *BRCA1* and *BRCA2* mutation carriers. We have now completed a large-scale genotyping study of 798 breast cancer cases and 840 controls from the Mayo Clinic for polymorphisms in genes encoding regulators of mitosis. The Mayo Clinic Breast Cancer study is an on-going clinic-based case-control study initiated in February 2001 at Mayo Clinic, Rochester, MN. Details of the study design and data collection procedures have been previously described (14) wrong reference. Briefly, cases were women over age 20 years with histologically confirmed primary invasive breast carcinoma who were enrolled within six months of date of diagnosis. Controls without prior history of cancer (other than non-melanoma skin cancer) were matched on age (± 5 years) and region of residence to cases. Controls were selected from the outpatient clinic in the Department of Internal Medicine at Mayo Clinic where they were seen for general medical examinations. Written informed consent was obtained from all participants. Case participation was 69% and control participation was 71%. The present analysis genotyped Caucasian women (99% of study participants) enrolled through June 30, 2005, representing 798 cases and 843 controls. Both the cases and controls completed a self-administered risk factor questionnaire. The questionnaire asked about known or suspected breast cancer risk factors including lifestyle, medical and reproductive factors. Cases and controls provided blood samples from which genomic DNA was isolated using standard protocols. The samples were bar coded to ensure accurate and reliable sample processing and storage.

This study entailed screening the dbSNP, HapMap, Perlegen, Seattle SNPs, and EGP websites for polymorphisms in 273 different mitotic genes, downloading genotyping data and selecting tagged SNPs based on the LD Select program with a minor allele frequency (MAF) >0.05 and an r^2 >0.80. Coding SNPs were selected based on a change in amino acid and a MAF >0.05. SNPs located in promoter regions and 5' and 3' UTRs were also included. A total of 2,400 SNPs from the 273 genes were selected and genotyped on the cases and controls along with 5% duplicate samples. Genotyping was highly successful. Only 165 SNPs failed genotyping. Of the remainder, call rates for genotypes were greater than 98%. Duplicates demonstrated 100% concordance. Only two SNPs were not in Hardy Weinberg Equilibrium (HWE) ($p<0.05$).

Individual SNP associations for breast cancer risk were assessed using unconditional logistic regression to estimate ORs and 95% CIs. Primary tests for associations were carried out assuming an ordinal (log-additive or additive) genotypic relationship using simple tests for trend within the logistic and linear regression models. All analyses were adjusted for the design variables of age and region of residence. We also examined the influence of demographic or clinical variables and excluded those variables that were not statistically significant at $p > 0.10$ using a backward elimination selection approach, performed separately for risk and density analyses. A

total of 144 SNPs displayed significant association with breast cancer risk. When accounting for correlations between SNPs in the same genes by random permutations of cases and controls we found that these positive associations represent a 30% increase over the number of associations expected by chance alone. Thus, several SNPs in regulators of cell division may influence the risk of breast cancer in the population.

We subsequently proposed to extend these findings into the *BRCA1* and BRCA2 carrier population. Specifically, we initiated a study aimed at evaluating the 144 SNPs from genes involved in regulation of cell division as modifiers of breast cancer risk in mutation carriers. We selected the Illumina 384-SNP goldengate array as the most cost-effective, high-quality genotyping platform for this study. This system is available in the genotyping core of the Mayo Clinic. To make full use of the 384 SNP array we selected the 144 SNPs and also selected SNPs from cell division control genes that commonly displayed associations with breast cancer risk in two breast cancer genome wide association studies conducted by Douglas Easton from Cambridge University (Easton et al., 2007) and from the CGEMS group at NCI (Hunter et al., 2007). Through our collaboration with Dr. Easton, we have access to genotyping data, odds ratios and p-values for all 12,026 SNPs that were evaluated in stage 2 of the genome wide study (Easton et al., 2007). In terms of CGEMS, odds ratios and p-values for all SNPs in Stage 1 of the Genome wide study are publicly available.

We are in the process of ordering SNPs in Stage 2 of the Easton study and Stage 1 of CGEMS by strength of association (p-value). All SNPs that reached both a significance of $p<0.001$ (approximately 700) in the Easton study and a significance of $p<0.01$ in CGEMS (5500) were selected. Likewise, all SNPs that reached both a significance of $p<0.001$ in CGEMS (550) and a significance of $p<0.01$ in the Easton study were selected (7000). It is important to note that these studies were performed on different platforms and that as a result many of the tagging SNPs from the same genes that displayed associations with risk on the two platforms were not the same. Thus, we mapped the SNPs displaying significant association with risk in either study into specific haplotype blocks ($r^2>0.6$) defined by HapMap data. Where SNPs from both studies mapped to a haplotype block, the association was considered validated and the SNP in the haplotype block displaying the most significant association with risk was selected. Likewise when several SNPs from a haplotype block exhibited significant association with risk only the SNP with the most significant association was selected. All of these SNPs were assessed for assay conversion on the Illumina Goldengate system through consultation with the Illumina Bioinformatics Support Center. The resulting list of SNPs that could be genotyped on the Goldengate platform were ordered by the significance of the association with risk. Those SNPs present in genes associated with cell division were then selected for genotyping and were combined with the initial 144 SNPs from our earlier study until 384 SNPs had been selected for the array study. A sentrix 384-bead array containing these 384 SNPS will shortly be ordered from Illumina Inc.

In parallel, we requested DNA samples and risk factor data from six collaborating groups. All have agreed to participate and are currently selecting and aliquoting these DNA samples for shipment to the Mayo Clinic. Epidemiological risk factor data matching all of these specimens are available through the CIMBA consortium database (Couch et al., 2007; Chenevix-Trench et al., 2007). A summary of the contributions of *BRCA1*/2 carriers from the major collaborating centers is as follows: 500 carriers will come from two collections at the Mayo Clinic (Drs. Couch and Szabo, PIs); 1200 will come from EMBRACE, a UK collection of carriers directed by Dr. Easton; 500 will come from GEO-HEBON, a Dutch national collection (Drs. Hogervorst and Rookus, PIs); 700 will come from Vienna (Dr. Furhauser, PI), 500 will come from the University of Pennsylvania (Dr. Nathanson, PI), 500 from a German National consortium managed by Dr. Schmutzler; 800 from Australia (Dr. Spurdle, PI). Only *BRCA1* or *BRCA2* female breast cancer cases or unaffected individuals are included. A total of 2,000 *BRCA1* mutation carriers (1,000 affected with breast cancer and 1,000 unaffected) and 2,000 *BRCA2* mutation carriers (1,000 affected with breast cancer and 1,000 unaffected) will be used for genotyping. We computed the statistical power of the study to account for multiple testing at a significance level of 0.05/384 ($\sim 10^{-4}$). At this level of significance and when genotyping 4,000 carriers the study has 80% power to detect a risk ratio of 1.3 for a SNP of MAF>0.20. Thus, the study is adequately powered to detect associations with small effect sizes.

Once these samples arrive (October 2007), they will be aliquoted at 50ng/μl into 96 well plates with 2%

duplicates and four controls (one water control and a CEPH trio). The samples will then be genotyped on the 384-array in the Mayo Clinic genotyping center. Genotyping data will be assessed for samples and SNP call rates. SNPs displaying call rates <95% will be excluded. Tests of Hardy Weinberg equilibrium (HWE) for *BRCA1* carriers will be performed and SNPs with $p<0.001$ will be excluded. Our primary analysis will be to evaluate the association of each SNP individually with breast cancer using a weighted Cox proportional hazards model that measures time to disease diagnosis and incorporates information on both disease status and age. Analyses will be adjusted for study center and geographical region to allow for differences in disease risks and allele frequencies. A robust variance estimation approach will be used to allow for more than one carrier from the same family. We will account for prophylactic oophorectomy in the analyses because oophorectomy is known to reduce risk of breast cancer by up to 50% in *BRCA1* and *BRCA2* mutation carriers (Rebbeck et al., 2002). Adjustment for other risk factors will also be performed.

At the conclusion of the study, in early to mid 2008, we expect to have identified a number of novel modifiers of breast cancer risk in *BRCA1* and *BRCA2* carriers. These modifiers will prove useful for identifying carriers who are at lower risk of cancer compared to all carriers and may benefit from a watchful waiting approach to cancer prevention as opposed to invasive prophylactic oophorectomy and mastectomy.

Aim 2. To demonstrate that Val57Ile alters STK15 function and co-operates with BRCA1/2 mutations to disrupt mitotic regulation.

Task 5 ands 6. When we began this study, the F31I variant had already been shown to alter the activity of the Aurora box-1 of STK15 protein, resulting in disruption of p53 binding and a decreased rate of degradation of STK15 (Ewart-Toland et al., 2003). It had also been shown by others that stabilized STK15 was associated with centrosome amplification and failure of cytokinesis, increased chromosomal instability and aneuploidy, suggesting a direct effect of the F31I variant on promotion of tumor formation (Ewart-Toland et al., 2003). As a result, Tasks 5 and 6 were deemed to be complete. Our subsequent finding that F31I does not influence breast cancer risk in *BRCA1* and *BRCA2* carriers suggests that these effects of STK15 stability make no contribution to cancer risk.

Aim 3. To establish the involvement of STK15 in breast tumor formation using Val57 and Ile57-STK15 transgenic mice and to evaluate synergism with BRCA1/2 by intercrossing with conditional brca1 and brca2 mutant mouse models.

Tasks 7-9. As noted above, neither STK15 F31I or V57I are associated with increased risk of breast cancer. On the basis of this finding we felt that it was inappropriate to continue with the proposed generation of transgenic animals expressing these mutant forms of STK15 in order to assess their infleunce on breast cancer development *in vivo*. Instead, we focused our efforts on Task 3 and 4 in an effort to identify variants in other mitotic regulators that modify the risk of breast cancer in *BRCA1* and BRCA2 carriers.

Key Research Accomplishments

- The F31I and V57I polymorphisms in STK15 are not associated with modification of breast cancer risk in *BRCA1* and *BRCA2* carriers.
- The F31I polymorphism in STK15 is not associated with breast cancer risk in a series of case-control studies.
- Common genetic variants in genes encoding mitotic regulators are associated with altered risk of breast cancer in a breast cancer case-control study.

Reportable Outcomes

1. The Breast Cancer Association Consortium. Commonly studied single nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. JNCI. 98:1382-1396, 2006.
2. Chenevix-Trench G, Milne RL, Antoniou AC, **Couch FJ**, Easton DF, Goldgar DE; CIMBA. An international initiative to identify genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers:

the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA). Breast Cancer Res. 9:104 [Epub], 2007.
3. Couch FJ, Sinilnikova O, Vierkant RA, Pankratz VS, Fredericksen ZS, Stoppa-Lyonnet D, Coupier I, Hughes D, Hardouin A, Berthet P, Peock S, Cook M, Baynes C, Hodgson S, Morrison PJ, Porteous ME, Jakubowska A, Lubinski J, Gronwald J, Spurdle AB; kConFab, Schmutzler R, Versmold B, Engel C, Meindl A, Sutter C, Horst J, Schaefer D, Offit K, Kirchhoff T, Andrulis IL, Ilyushik E, Glendon G, Devilee P, Vreeswijk MP, Vasen HF, Borg A, Backenhorn K, Struewing JP, Greene MH, Neuhausen SL, Rebbeck TR, Nathanson K, Domchek S, Wagner T, Garber JE, Szabo C, Zikan M, Foretova L, Olson JE, Sellers TA, Lindor N, Nevanlinna H, Tommiska J, Aittomaki K, Hamann U, Rashid MU, Torres D, Simard J, Durocher F, Guenard F, Lynch HT, Isaacs C, Weitzel J, Olopade OI, Narod S, Daly MB, Godwin AK, Tomlinson G, Easton DF, Chenevix-Trench G, Antoniou AC; on behalf of the Consortium of Investigators of Modifiers of BRCA1/2. *AURKA* F31I Polymorphism and Breast Cancer Risk in *BRCA1* and *BRCA2* Mutation Carriers: the Consortium of Investigators of Modifiers of *BRCA1* and BRCA2 (CIMBA). Cancer Epi Bio & Prev. 16:1416-1421, 2007.

Conclusions

We have used very large datasets to demonstrate that the F31I polymorphism in STK15 does not increase the risk of breast cancer in carriers of *BRCA1* and *BRCA2* mutations. A similar lack of effect was seen in another very large pooled dataset from sporadic breast cancer case-control studies. However, it is likely that polymorphisms in other mitotic regulators alter breast cancer risk in sporadic and familial breast cancer patients. We have evaluated a number of such variants in a breast cancer case-control study and have initiated a study aimed at validating these findings in *BRCA1* and *BRCA2* carriers. Specifically, we are gathering DNA and risk factor data from 4,000 *BRCA1* and *BRCA2* mutation carriers from six collaborating centers. Once these samples are in hand (expected by October 2007), the variants will be genotyped on a 384-SNP Goldgengate array and assessed for breast cancer risk modification in *BRCA1* and *BRCA2* carriers.

This work (Tasks 3 and 4) is not complete so we have filed for a no-cost extension of the project until 5-20-2008.

References

Fodor FH, Weston A, et al. (1998). "Frequency and carrier risk associated with common BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer patients." Am J Hum Genet. **63**(1):45-51.

Goldgar DE, Fields P, et al. (1994). "A large kindred with 17q-linked breast and ovarian cancer: genetic, phenotypic, and genealogical analysis." J Natl Cancer Inst. **86**(3):200-9.

Rebbeck TR.(2002). "Inherited predisposition and breast cancer: modifiers of BRCA1/2-associated breast cancer risk." Environ Mol Mutagen. **39**(2-3):228-34.

Gayther SA, Warren W, et al. (1995). "Germline mutations of the BRCA1 gene in breast and ovarian cancer families provide evidence for a genotype-phenotype correlation." Nat Genet. **11**(4): 428-33.

Gayther SA, Mangion J, et al. (1997). "Variation of risks of breast and ovarian cancer associated with different germline mutations of the BRCA2 gene." Nat Genet. **15**(1): 103-5.

Rebbeck TR. (1999). "Inherited genetic predisposition in breast cancer. A population-based perspective." Cancer. **86**(11 Suppl): 2493-501.

Hsu LC, Doan TP, et al. (2001). "Identification of a gamma-Tubulin-binding Domain in BRCA1." Cancer Res. **61**(21): 7713-8.

Nakanishi A, Han X, et al. (2007). Interference with BRCA2, which localizes to the centrosome during S and early M phase, leads to abnormal nuclear division. Biochem Biophys Res Commun. 355(1):34-40.

Xu X, Weaver Z, et al. (1999). “Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells.” Mol Cell. 3(3): 389-95.

Deng CX. (2002). “Roles of BRCA1 in centrosome duplication.” Oncogene. 21(40):6222-7.

Starita LM, Machida Y, et al. (2004). “BRCA1-dependent ubiquitination of gamma-tubulin regulates centrosome number”. Mol Cell Biol. 24(19):8457-66.

Wu K, Hinson SR, et al. (2005). “Functional evaluation and cancer risk assessment of BRCA2 unclassified variants”. Cancer Res. 65(2):417-6.

Ouchi M, Fujiuchi N, et al. (2004). “BRCA1 phosphorylation by Aurora-A in the regulation of G2 to M transition.” J Biol Chem. 279(19):19643-8.

Ewart-Toland A, Briassouli P, et al. (2003) “Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human.” Nat Genet. 34(4):403-12.

Egan KM, Newcomb PA, et al. (2004). “STK15 polymorphism and breast cancer risk in a population based study”. Carcinogenesis 25:2149-53.

Sun T, Miao X, et al. (2004). “Functional Phe31Ile polymorphism in Aurora A and risk of breast carcinoma”. Carcinogenesis 25: 2225-30.

Lo YL, Yu JC, et al. (2005). “Breast cancer risk associated with genotypic polymorphism of the mitosis-regulating gene Aurora-A/STK15/BTAK”. Int J Cancer 115:276-83.

Ewart-Toland A, Dai Q, et al. (2005). « Aurora-A/STK15 T+91A is a general low penetrance cancer susceptibility gene: a meta-analysis of multiple cancer types”. Carcinogenesis 26:1368-73.

Cox DG, Hankinson SE and Hunter DJ. (2006). “Polymorphisms of the AURKA (STK15/Aurora Kinase) Gene and Breast Cancer Risk” Cancer Causes & Control 17:81-3.

Dai Q, Cai QY, et al. (2004). “Synergistic effects of STK15 gene polymorphisms and endogenous estrogen exposure in the risk of breast cancer”. Cancer Epidemiol Biomarkers Prev 13:2065-70.

Fletcher O, Johnson N, et al. (2006). “Inconsistent association between the STK15 F31I genetic polymorphism and breast cancer risk”. J Natl Cancer Inst 98:1014-18.

Rebbeck TR, Lynch HT, et al. (2002). “Prevention and Observation of Surgical End Points Study Group. Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations”. N Engl J Med 346:1616-22.

Couch FJ, Sinilnikova O, et al. (2007). “*AURKA* F31I Polymorphism and Breast Cancer Risk in *BRCA1* and *BRCA2* Mutation Carriers: the Consortium of Investigators of Modifiers of *BRCA1* and BRCA2 (CIMBA)”. Cancer Epi Bio & Prev. 16:1416-1421, 2007.

Easton DF, Pooley KA, et al. (2007). “Genome-wide association study identifies novel breast cancer susceptibility loci”. Nature. 447(7148):1087-93.

Hunter DJ, Kraft P, et al. (2007). “A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer”. Nat Genet. 39(7):870-874.

AURKA F31I Polymorphism and Breast Cancer Risk in *BRCA1* and *BRCA2* Mutation Carriers: A Consortium of Investigators of Modifiers of BRCA1/2 Study

Fergus J. Couch,[1] Olga Sinilnikova,[4] Robert A. Vierkant,[3] V. Shane Pankratz,[3] Zachary S. Fredericksen,[3] Dominique Stoppa-Lyonnet,[6] Isabelle Coupier,[6] David Hughes,[5] Agnès Hardouin,[7] Pascaline Berthet,[7] Susan Peock,[8] Margaret Cook,[8] Caroline Baynes,[9] Shirley Hodgson,[10] Patrick J. Morrison,[11] Mary E. Porteous,[12] Anna Jakubowska,[13] Jan Lubinski,[13] Jacek Gronwald,[13] Amanda B. Spurdle,[14] kConFab,[15] Rita Schmutzler,[16] Beatrix Versmold,[16] Christoph Engel,[17] Alfons Meindl,[18] Christian Sutter,[19] Jurgen Horst,[21] Dieter Schaefer,[22] Kenneth Offit,[23] Tomas Kirchhoff,[23] Irene L. Andrulis,[24,25] Eduard Ilyushik,[24] Gordon Glendon,[25] Peter Devilee,[27] Maaike P.G. Vreeswijk,[27] Hans F.A. Vasen,[28] Ake Borg,[29] Katja Backenhorn,[29] Jeffery P. Struewing,[30] Mark H. Greene,[31] Susan L. Neuhausen,[32] Timothy R. Rebbeck,[33] Katherine Nathanson,[33] Susan Domchek,[33] Theresa Wagner,[35] Judy E. Garber,[36] Csilla Szabo,[1] Michal Zikan,[37] Lenka Foretova,[38] Janet E. Olson,[3] Thomas A. Sellers,[39] Noralane Lindor,[2] Heli Nevanlinna,[40] Johanna Tommiska,[40] Kristiina Aittomaki,[41] Ute Hamann,[20] Muhammad U. Rashid,[20] Diana Torres,[20] Jacques Simard,[42] Francine Durocher,[42] Frederic Guenard,[42] Henry T. Lynch,[43] Claudine Isaacs,[44] Jeffrey Weitzel,[45] Olufunmilayo I. Olopade,[46] Steven Narod,[26] Mary B. Daly,[34] Andrew K. Godwin,[34] Gail Tomlinson,[47] Douglas F. Easton,[8] Georgia Chenevix-Trench,[14] and Antonis C. Antoniou[8] on behalf of the Consortium of Investigators of Modifiers of BRCA1/2

[1]Departments of Laboratory Medicine and Pathology, [2]Medical Genetics, and [3]Health Sciences Research, Mayo Clinic College of Medicine, Rochester, Minnesota; [4]Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon-Centre Léon Bérard; [5]IARC, Lyon, France; [6]Institut National de la Sante et de la Recherche Medicale U509, Service de Génétique Oncologique, Institut Curie, Paris, France; [7]Centre François Baclesse, Caen, France; [8]Cancer Research UK, Genetic Epidemiology Unit, Department of Public Health and Primary Care and [9]Cancer Research UK, Human Cancer Genetics Group, Department of Oncology, University of Cambridge, Cambridge, United Kingdom; [10]South West Thames Regional Genetics Service, St. Georges Hospital, London, United Kingdom; [11]Northern Ireland Regional Genetics Centre, Cancer Genetics Service, Belfast City Hospital, Belfast, United Kingdom; [12]South East of Scotland Clinical Genetics Service, Western General Hospital, Edinburgh, United Kingdom; [13]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland; [14]Queensland Institute of Medical Research, Brisbane, Queensland, Australia; [15]The Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia; [16]Division of Molecular Gyneco-Oncology, University of Cologne, Cologne, Germany; [17]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany; [18]Department of Gynaecology and Obstetrics, Technical University, Munich, Germany; [19]Institute of Human Genetics, University of Heidelberg; [20]Deutsches Krebsforschungszentrum Heidelberg, Heidelberg, Germany; [21]Institute of Human Genetics, University of Muenster, Muenster, Germany; [22]Institute of Human Genetics, University of Frankfurt, Frankfurt, Germany; [23]Clinical Genetics Service, Memorial Sloan-Kettering Cancer Center, New York, New York; [24]Fred A. Litwin Center for Cancer Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital; [25]Ontario Cancer Genetics Network, Cancer Care Ontario; [26]Centre for Research in Women's Health, Toronto, Ontario, Canada; [27]Department of Human Genetics, Leiden University Medical Center; [28]The Foundation for the Detection of Hereditary Tumors, Leiden, the Netherlands; [29]Department of Oncology, Lund University Hospital, Lund, Sweden; [30]Clinical Genetics Branch and [31]Laboratory of Population Genetics, National Cancer Institute, NIH, Bethesda, Maryland; [32]Division of Epidemiology, Department of Medicine, University of California Irvine, Irvine, California; [33]Abramson Cancer Center, University of Pennsylvania School of Medicine; [34]Fox Chase Cancer Center, Philadelphia, Pennsylvania; [35]Division of Senology, Medical University of Vienna, Vienna, Austria; [36]Dana-Farber Cancer Institute, Boston, Massachusetts; [37]Department of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, Prague, Czech Republic; [38]Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic; [39]Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida; Departments of [40]Obstetrics and Gynecology and [41]Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland; [42]Cancer Genomics Laboratory, Oncology and Molecular Endocrinology Research Centre, Centre Hospitalier Universitaire de Québec and Laval University, Quebec City, Quebec, Canada; [43]Department of Preventive Medicine and Public Health, Creighton University, Omaha, Nebraska; [44]Lombardi Cancer Center, Georgetown University, Washington, District of Columbia; [45]City of Hope Cancer Center, Duarte, California; [46]University of Chicago, Chicago, Illinois; and [47]University of Texas Southwestern Medical Center at Dallas, Dallas, Texas

Abstract

The AURKA oncogene is associated with abnormal chromosome segregation and aneuploidy and predisposition to cancer. Amplification of AURKA has been detected at higher frequency in tumors from BRCA1 and BRCA2 mutation carriers than in sporadic breast tumors, suggesting that overexpression of AURKA and inactivation of BRCA1 and BRCA2 cooperate during tumor development and progression. The F31I polymorphism in AURKA has been associated

Received 2/19/07; revised 4/10/07; accepted 4/23/07.

Grant support: Breast Cancer Research Foundation, U.S. Army Medical Research and Materiel Command grant W81XWH-04-1-0588, and Mayo Clinic Breast Cancer Specialized Program of Research Excellence grant P50-CA116201 (F.J. Couch); Cancer Care Ontario and RFA CA-95-003 as part of the National Cancer Institute Breast Cancer Family Registries (I.L. Andrulis); Canadian Institute for Health Research through the INHERIT BRCAs program (J. Simard); Dutch Cancer Society grant UL2001-2471 (P. Devilee); Academy of Finland (110663), Finnish Cancer Society, Helsinki University Central Hospital Research Fund, and Sigrid Juselius Fund (H. Nevanlinna); Deutsches Krebsforschungszentrum Heidelberg (U. Hamann); NIH grants R01-CA74415 (S.L. Neuhausen) and R01-CA083855 and R01-CA102776 (T.R. Rebbeck); The Programme Hospitalier de Recherche Clinique AOR01082 (D. Stoppa-Lyonnet and O. Sinilnikova); Komen Foundation BCTR 0601361 and Weissenbach-Southworth-Niehaus Research Fund (K. Offitt); The German Cancer Aid grant 107054 and Center for Molecular Medicine Cologne grant TV 93 (R. Schmutzler). A General Clinic Research Center grant from the NIH (M01 RR00043) awarded to the City of Hope National Center supports in part the collection and management of the Hereditary Cancer Registry (J. Weitzel). kConFab is supported by grants from the National Breast Cancer Foundation, the National Health and Medical Research Council, and by the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania, and South Australia, and the Cancer Foundation of Western Australia. D.F. Easton is a principal research fellow of Cancer Research UK. EMBRACE and A. Antoniou are funded by an award from Cancer Research UK.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Fergus J. Couch, Mayo Clinic College of Medicine, 200 First Street Southwest, Rochester, MN 55905. Phone: 507-284-3623; Fax: 507-266-0824. E-mail: Couch.fergus@mayo.edu

doi:10.1158/1055-9965.EPI-07-0129

with breast cancer risk in the homozygous state in prior studies. We evaluated whether the AURKA F31I polymorphism modifies breast cancer risk in BRCA1 and BRCA2 mutation carriers from the Consortium of Investigators of Modifiers of BRCA1/2. Consortium of Investigators of Modifiers of BRCA1/2 was established to provide sufficient statistical power through increased numbers of mutation carriers to identify polymorphisms that act as modifiers of cancer risk and can refine breast cancer risk estimates in BRCA1 and BRCA2 mutation carriers. A total of 4,935 BRCA1 and 2,241 BRCA2 mutation carriers and 11 individuals carrying both BRCA1 and BRCA2 mutations was genotyped for F31I. Overall, homozygosity for the 31I allele was not significantly associated with breast cancer risk in BRCA1 and BRCA2 carriers combined [hazard ratio (HR), 0.91; 95% confidence interval (95% CI), 0.77-1.06]. Similarly, no significant association was seen in BRCA1 (HR, 0.90; 95% CI, 0.75-1.08) or BRCA2 carriers (HR, 0.93; 95% CI, 0.67-1.29) or when assessing the modifying effects of either bilateral prophylactic oophorectomy or menopausal status of BRCA1 and BRCA2 carriers. In summary, the F31I polymorphism in AURKA is not associated with a modified risk of breast cancer in BRCA1 and BRCA2 carriers. (Cancer Epidemiol Biomarkers Prev 2007;16(7):1416–21)

Introduction

The *AURORA-A/AURKA/BTAK/STK15* gene encodes a serine/threonine kinase that regulates mitotic chromosome segregation. *AURKA* is amplified and overexpressed in breast and other tumors and is associated with centrosome amplification, failure of cytokinesis, and aneuploidy. Genetic mapping studies in mouse models suggest that *AURKA* is a genetic modifier of cancer risk (1). In addition, mouse models of *AUR7KA* exhibit infrequent mammary gland tumor formation but display synergy in tumor formation when combined with overexpressed oncogenes or disrupted tumor suppressors, suggesting that *AURKA* is a low-risk cancer susceptibility gene (2).

Further evidence for a role of *AURKA* in breast cancer comes from observations that homozygosity for a F31I polymorphism in *AURKA* is associated with an increased risk for breast cancer. In a study of incident breast cancer cases (n = 941) and age-matched population controls (n = 830), Egan et al. (3) found that the breast cancer risk for Ile/Ile homozygotes were at increased risk for breast cancer [odds ratio (OR), 1.54; 95% confidence interval (95% CI), 0.96-2.47], although this finding was not significant. Sun et al. (4) observed that the Ile-encoding allele is the common allele in the Chinese population, whereas the Phe-encoding allele is more common in Caucasian populations (4). In addition, an association between Ile/Ile homozygotes and estrogen receptor–negative breast carcinomas (OR, 2.56; 95% CI, 1.24-5.26) was detected. Lo et al. (5) reported a significant association between *AURKA* haplotypes and breast cancer risk. Ewart-Toland et al. (6) also found an increase in cancer risk for the *Ile/Ile* homozygotes (OR, 1.35; 95% CI, 1.12-1.64; P = 0.002) in a meta-analysis of data from four case-control breast cancer populations. Furthermore, postmenopausal women homozygous for the F31I and I57V alleles of *AURKA* in a case-control study nested within the Nurses' Health Study prospective cohort had an increased risk of invasive breast cancer (OR, 1.63; 95% CI, 1.08-2.45; ref. 7). In contrast, Dai et al. (8) did not observe a significant association with breast cancer risk for Ile/Ile homozygotes (OR, 1.2; 95% CI, 0.9-1.6) in a population-based case-control series of Han Chinese, and Fletcher et al. (9) found no association between Ile/Ile homozygotes and risk of bilateral breast cancer (OR, 0.63; 95% CI, 0.34-1.13). Importantly, the F31I variant has been shown to alter the activity of the Aurora box-1 of the AURKA protein, resulting in disruption of p53 binding and a decreased rate of degradation of AURKA. The stabilized AURKA may lead to centrosome amplification and failure of cytokinesis, increased chromosomal instability and aneuploidy, and promotion of tumor formation (1).

Mutations in *BRCA1* and *BRCA2* are correlated with aberrant duplication of the centrosome leading to centrosome amplification, chromosome missegregation, and aneuploidy (10-12). Amplification of *AURKA* has also been detected at much higher frequency in tumors from *BRCA1* and *BRCA2* mutation carriers than in sporadic breast tumors, suggesting that overexpression of AURKA and inactivation of BRCA1 and BRCA2 cooperate during tumor development and/or progression. Based on these data, we hypothesized that the F31I polymorphism modifies the risk of breast cancer in *BRCA1* and *BRCA2* mutation carriers. To address this hypothesis, *AURKA* F31I was genotyped on *BRCA1* and *BRCA2* deleterious mutation carriers from 16 clinic and population-based research studies and multicenter consortia participating in the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) and the association of F31I with breast cancer risk was assessed.

Materials and Methods

Subjects. *BRCA1* and *BRCA2* mutation carriers were identified through 16 clinic and population-based research studies and multicenter consortia participating in the CIMBA. This international consortium was established in 2005 by a group of investigators interested in identifying modifiers of cancer risk in *BRCA1* and *BRCA2* mutation carriers that could be used to refine cancer risk estimates. Recruitment of mutation carriers for this and other CIMBA studies was approved by institutional review boards or ethics committees at all sites. *BRCA1* and *BRCA2* mutation carriers were defined as carriers of frameshifting small deletions and insertions, nonsense mutations, splice site mutations verified *in vitro*, and large genomic rearrangements that result in a premature stop codon in either *BRCA1* or *BRCA2*. These mutations were identified by a variety of screening techniques and sequence verified. As the K3326X variant in exon 27 is not associated with high risk of breast cancer, this and other mutations causing stop codons in exon 27 were excluded. Missense mutations that have been classified as pathogenic by multifactorial likelihood approaches were included in the deleterious category (12-14), whereas carriers of all other missense and intronic mutations in *BRCA1* and *BRCA2* were excluded from the study. Phenotypic data for mutation carriers were provided by each contributing center. Data were collected on year of birth, mutation description, ethnicity, country of residence, age at last follow-up, ages at breast and ovarian cancer diagnosis, age at bilateral prophylactic mastectomy, age at bilateral prophylactic oophorectomy, and status and age at menopause. These and other available epidemiologic data obtained from risk factor questionnaires and/or medical records were uniformly coded and stored in a centralized CIMBA database.

Genotyping. The F31I polymorphism (rs2273535) of *AURKA* was genotyped by 13 groups by the 5′ nuclease assay (Taqman) on an ABI 7900HT Sequence Detection System (Applied Biosystems). PCR primers were 5′-CTGGCCACTATTTACAGGTAATGGA-3′ (forward) and 5′-TGGAGGTCCAAAACGTGTTCTC-3′ (reverse). Probes were VIC-ACTCAGCAATTTCCTT and FAM-CTCAGCAAATTCCTT. The annealing temperature was 60°C. Lund investigators used an alternative reverse primer (CATCTTTTGCTTTCATGAATGCCAG) and did the 5′ nuclease assay on a RotorGene

Table 1. Characteristics of study subjects by site

Source	Ascertainment	*BRCA1* cases	*BRCA1* unaff.*	Total *BRCA1*	*BRCA2* cases	*BRCA2* unaff.	Total *BRCA2*	*B1/2* † cases	*B1/2* unaff.	Total *B1/2*	Total carriers
MAGIC	Clinic	303	428	731	137	160	297	3	0	3	1,031
GEMO	Clinic	413	276	689	223	84	307	0	0	0	996
EMBRACE	Clinic	235	219	454	156	148	304	1	2	3	761
Poland	Clinic	307	427	734	0	0	0	0	0	0	734
kConFab	Clinic	203	201	404	169	143	312	0	0	0	716
GCHBOC	Clinic	286	113	399	173	52	225	3	0	3	627
MSKCC	Clinic	174	117	291	102	70	172	1	0	1	464
Ontario	Clinic and population	125	52	177	100	41	141	0	0	0	318
LUMC	Clinic	99	120	219	12	20	32	0	0	0	251
Lund	Clinic	73	88	161	38	32	70	0	0	0	231
MOD-SQUAD	Clinic	82	67	149	28	15	43	0	0	0	192
HEBCS	Clinic	56	39	95	54	40	94	0	0	0	189
DKFZ	Clinic	82	41	123	30	21	51	0	0	0	174
MAYO	Clinic	53	23	76	26	20	46	0	0	0	122
INHERIT	Clinic	33	37	70	40	41	81	0	0	0	151
NCI	Clinic	47	116	163	17	50	67	0	0	0	230
Total		2,571	2,364	4,935	1,305	937	2,242	8	2	10	7,187

Abbreviations: MAGIC, Modifiers and Genetics in Cancer; GEMO, Genetic Modifiers of cancer risk in BRCA1/2 mutation carriers study; GCHBOC, German Consortium for Hereditary Breast and Ovarian Cancer; EMBRACE, Epidemiological Study of *BRCA1* and BRCA2 Mutation Carriers; kConFab, Kathleen Cunningham Consortium for Research into Familial Breast Cancer; INHERIT BRCAs, Interdisciplinary Health Research International Team on Breast Cancer susceptibility; MSKCC, Memorial Sloan-Kettering Cancer Center; MAYO, Mayo Clinic; LUMC, Leiden University Medical Center; MOD-SQUAD, Modifier Study of Quantitative Effects on Disease; HEBCS, Helsinki Breast Cancer Study; DKFZ, Deutsches Krebsforschungszentrum Heidelberg; NCI, National Cancer Institute.
*The term unaff. refers to individuals not affected with breast cancer.
†*B1/2* refers to individuals with both *BRCA1* and *BRCA2* deleterious mutations.

(Corbett Research). INHERIT investigators directly sequenced the polymorphism using the following primers: 5′-GGGTG-AGGAATTGGAGGGGAT-3′ (forward) and 5′-GGACACCA-ATTTATGCTGTGTCCT-3′ (reverse). Genotyping for the HEBCS was done by Amplifluor fluorescent genotyping (KBioscience).[48] Genotyping for the DKFZ and Polish studies was done by fragment analysis. DNA fragments containing the polymorphism were amplified using forward primer 5′-AGTTGGAGGTCCAAAACGTG-3′ and Cy5-labeled reverse primer 5′-CGCTGGGAAGTATTTGAAGG-3′, digested with 2.5 units *Xap*I (Fermentas), separated on 3% agarose gel (Polish samples) or by capillary gel electrophoresis (German samples) on a CEQ 8000 DNA Analysis System (Beckmann), and sized relative to CEQ DNA Size Standard-400 in each well. Allele sizes were 114 bp for the T allele and 78 bp for the A allele.

Statistical Methods. Hazard ratios (HR) were modeled using Cox proportional hazards regression analysis, with breast cancer as the outcome and age as the time variable (15). We corrected for possible ascertainment bias using a weighted cohort approach (16). Briefly, this involves assigning weights to the mutation-carrying subjects such that the reweighted incidence rates observed in the study sample are consistent with the age-dependent penetrances for breast cancer onset established in carriers of inactivating mutations in BRCA1 and BRCA2. Subjects were followed from birth until the earliest occurrence of breast cancer (3,884), bilateral prophylactic mastectomy (232), ovarian cancer (643), age 80 (97), or age at last contact (2,331). Subjects were censored at age 80 because population-based incidence rates for older mutation carriers are unreliable, and accurate sampling weights cannot be assigned. Carriers with both *BRCA1* and *BRCA2* mutations were included once in overall analyses and were also included in each of the *BRCA1* and *BRCA2* gene-specific analyses. The number of subjects in each family varied from 1 to 33, with 75% of families represented by a single individual. Because the exact relationships among the family members were not available, we accounted for the nonindependence of observations within families using a robust variance estimate (17). Primary analyses modeled *AURKA* as a recessive effect, comparing those with two copies of the minor allele with those with less than two copies. Secondary analyses examined associations using a two degree-of-freedom general model, simultaneously comparing subjects with one copy or with two copies of the minor allele with the subjects with zero copies.

Overall analyses were carried out for all subjects regardless of whether they carried a mutation in *BRCA1* or *BRCA2* or both. All analyses accounted for birth cohort and country of residence by including them as stratification variables in the Cox regression. The overall analysis also accounted for study site and mutation status. Additional analyses were conducted to obtain risk estimates for individuals with different characteristics, as defined by gene status, menopausal status, oophorectomy status, and study site. Gene-specific results accounted for study site along with birth cohort and country of residence by use of stratification variables. Site-specific results accounted for mutation status, birth cohort, and country of residence. Menopausal status and oophorectomy status were modeled as time-dependent covariates and results accounted for group status and mutation status. In secondary analyses, the influence of benign prophylactic oophorectomy and menopausal status on associations between the Ile/Ile genotype and breast cancer risk was also evaluated. As these covariates did not confound the observed associations, the associations reported in Table 2 are not adjusted for these variables.

Among those who provided ethnicity information, 97% were Caucasian, 2% were Ashkenazi Jewish, and the remaining 1% were ''other.'' Those who did not provide ethnicity information were grouped in a separate ''missing'' category for analysis purposes. Ethnicity was initially included as an additional stratification variable but was subsequently excluded because of the absence of any effect on the results. We assessed the possible heterogeneity of risk ratios across study site using standard tests of interaction. A sensitivity analysis assessing the effect of possible survival bias was conducted by excluding cases ascertained more than 3 years after diagnosis. All statistical tests were two sided, and all analyses were carried out using the Statistical Analysis System (SAS Institute, Inc.) and S-Plus (Insightful) software systems.

48http://www.kbioscience.co.uk

Results

A total of 4,935 female *BRCA1*, 2,241 female *BRCA2* deleterious mutation carriers, and 11 individuals carrying both *BRCA1* and *BRCA2* mutations was included in this study. Of these 7,187 mutation carriers, 3,884 had a diagnosis of breast cancer at the end of follow-up and 3,303 were censored as unaffected at a mean age of 43.4 years. The distribution of *BRCA1* and *BRCA2* carriers by study site, gene, and cancer status is shown in Table 1. To avoid overlap between studies, we compared carriers by country of origin, year of birth, mutation, and reported ages. Duplication of samples between MAYO and MAGIC and between GEMO and MAGIC was detected. In both instances, the duplicated samples were excluded from the MAGIC data set.

The distribution of the *AURKA* F31I genotypes is shown in Table 2. Of the 363 (5%) carriers homozygous for the Ile-encoding allele, 188 were affected with breast cancer. The frequency of the recessive Ile/Ile-encoding genotype in the 16 groups varied between 3% and 8%, which is similar to estimates from other populations (6). There was no difference in the frequency of the Ile/Ile recessive genotype across genotyping platforms ($P = 0.33$). Similarly, the study sites with the highest Ile/Ile frequencies did not have ethnic mixtures significantly different to the other study sites. The F31I polymorphism did not deviate significantly from Hardy-Weinberg equilibrium ($P = 0.07$) among all 7,187 affected and unaffected carriers.

The estimated risk of breast cancer associated with the recessive genotype for F31I in *BRCA1* and *BRCA2* carriers using a weighted Cox proportional hazards model is shown in Table 2. Although there was a suggestion of a protective effect (HR, 0.91; 95% CI, 0.77-1.06), overall, the result was not statistically significant. Similarly, no association with risk was observed for individual participating centers other than for two centers (Ontario and HEBCS) that contributed small numbers of carriers to the study (Table 2). A test for heterogeneity across study site was not significant ($P = 0.06$). In an effort to account for the trend toward heterogeneity, we investigated the influence of the three sites that were significantly different from the other sites [MOD-SQUAD ($P = 0.02$), GEMO ($P = 0.01$), and DKFZ ($P = 0.03$)] on the overall effect. Exclusion of each site in turn did not substantially alter the overall HR or the significance of the association.

Because BRCA1 is phosphorylated by AURKA (18), we evaluated whether the Ile/Ile genotype was associated with risk of breast cancer in *BRCA1* or *BRCA2* carriers. No significant association with risk was detected for either *BRCA1* (HR, 0.90; 95% CI, 0.75-1.08) or *BRCA2* carriers (HR, 0.93; 95% CI, 0.67-1.29; Table 2). As other studies have reported an association between the recessive Ile/Ile-encoding genotype and postmenopausal status in noncarriers (3, 7), we considered the influence of menopausal status of carriers on breast cancer risk. At the end of follow-up, 4,201 carriers were premenopausal and 2,986 were postmenopausal. No significant association with risk was detected (Table 2). Because prophylactic oophorectomy substantially reduces the risk of breast cancer in *BRCA1* and *BRCA2* mutation carriers (19), we also evaluated the influence of prophylactic oophorectomy status. A total of 707 individuals reported undergoing prophylactic oophorectomy, 4,298 reported no history of oophorectomy, whereas 2,182 (30%) provided no data at last follow-up. Associations with breast cancer risk by category of prophylactic oophorectomy did not differ markedly from the overall results. Secondary analyses using a two degree-of-freedom general model also failed to detect a significant association for either a single copy ($P = 0.97$) or two copies ($P = 0.24$) of the F31I polymorphism compared with no copies.

In an effort to account for possible survival bias and the inclusion of prevalent cases in the collection of *BRCA1* and

Table 2. Association of *AURKA* F31I with breast cancer risk

Group	0 or 1 copy Ile allele			2 copies Ile allele			HR (95% CI), all cases	HR (95% CI),* incident cases
	Unaffected	Affected	Person-years	Unaffected	Affected	Person-years		
Overall	3,128	3,696	296,122	175	188	15,793	0.91 (0.77-1.06)	0.84 (0.65-1.08)
By mutation status								
BRCA1	2,237	2,460	200,406	129	120	10,754	0.90 (0.75-1.08)	0.90 (0.66-1.22)
BRCA2	893	1,245	96,110	46	68	5,039	0.93 (0.67-1.29)	0.67 (0.44-1.03)
By menopausal status								
Premenopausal	1,935	2,049	242,208	111	106	12,834	0.84 (0.69-1.03)	0.83 (0.60-1.15)
Postmenopausal	1,193	1,647	53,914	64	82	2,959	0.96 (0.75-1.23)	0.77 (0.51-1.16)
By oophorectomy status								
No	1,772	2,318	201,303	101	107	10,474	0.85 (0.69-1.05)	0.82 (0.58-1.15)
Yes	510	160	3,793	28	9	213	1.10 (0.56-2.18)	1.03 (0.39-2.78)
Missing	846	1,218	91,026	46	72	5,106	0.97 (0.75-1.26)	0.86 (0.55-1.34)
By study site								
MAGIC	559	423	41,554	29	20	2,002	1.02 (0.63-1.67)	
GEMO	347	597	40,913	13	39	2,266	1.33 (0.97-1.82)	
EMBRACE	353	378	30,757	16	14	1,318	0.70 (0.37-1.32)	
Poland	399	285	30,360	28	22	2,197	0.98 (0.65-1.47)	
kConFab	322	362	29,568	22	10	1,251	0.64 (0.34-1.22)	
GCHBOC	157	432	24,819	8	30	1,698	0.94 (0.65-1.37)	
MSKCC	182	268	19,371	5	9	591	0.79 (0.38-1.66)	
Ontario	79	217	13,069	14	8	1,012	0.33 (0.13-0.82)	
LUMC	129	106	10,350	11	5	715	0.68 (0.32-1.44)	
Lund	113	102	11,401	7	9	803	1.05 (0.55-1.99)	
MOD-SQUAD	78	104	7,760	4	6	388	1.56 (1.04-2.36)	
HEBCS	75	108	8,451	4	2	344	0.27 (0.05-1.96)	
DKFZ	61	110	6,714	1	2	109	7.05 (0.66-75.2)	
MAYO	41	71	4,998	2	8	442	1.41 (0.65-3.07)	
INHERIT	76	70	6,668	2	3	225	1.29 (0.45-3.67)	
NCI	157	63	9,371	9	1	433	0.28 (0.05-1.77)	

NOTE: Weighted Cox proportional hazards regression analysis, modeling *AURKA* F31I as a recessive genotypic effect. Results overall by menopausal status and by oophorectomy status account for birth cohort, group status, country, and mutation status. Mutation-specific results account for birth cohort, group status, and country. Group-specific results account for birth cohort, mutation status, and country. Robust variance estimates were used to correct for possible nonindependence of study subjects.

*Cox proportional hazards regression analysis restricted to cases for whom genetic diagnosis is less than 3 y after breast cancer diagnosis.

BRCA2 carriers, we repeated our analysis after excluding cases diagnosed more than 3 years before the date of ascertainment. For this analysis, we excluded records where an age at interview was not provided. Overall, the mean difference between age of diagnosis and age at interview for the 3,422 cases with available data was 8.7 years. Of these, 1,322 (38.6%) cases had been diagnosed less than 3 years before the date of ascertainment. When excluding prevalent cases, no association between the Ile/Ile genotype and breast cancer risk was observed, and the risk estimates were similar to those obtained when using both prevalent and incident cases (Table 2).

Discussion

Overall, no evidence of a significant association between homozygosity for the F31I *AURKA* polymorphism and breast cancer risk in *BRCA1* and *BRCA2* mutation carriers in combination or alone was observed. These results were somewhat unexpected given the known functional relationship between AURKA and BRCA1 (18), the known influence of F31I on AURKA protein stability (1), and the significant associations with cancer risk reported in several studies of unselected breast cancer cases and controls. Although the variant does not seem to modify predisposition to cancer in this combined group of mutation carriers, the possibility remains that the Ile/Ile genotype influences tumor progression or clinical outcome or modifies cancer risk in conjunction with other risk factors. The suggestion of a modestly protective effect of the Ile/Ile genotype in this study particularly when restricting the study to incident cases supports this possibility. Interestingly, a study of bilateral breast cancer cases also identified a nonsignificant protective effect for the Ile/Ile genotype (9). This common protective effect among individuals at higher risk of breast cancer in the Caucasian population suggests that homozygosity for the F31I polymorphism may reduce cancer risk in high-risk groups while possibly increasing risk in the general population. Additional studies of other high-risk populations and the combined effects of other risk factors are needed to further evaluate these possibilities.

In this study, we accounted for the effects of both bilateral prophylactic oophorectomy and menopausal status effects by treating these factors as time-dependent variables in the analysis. As bilateral prophylactic oophorectomy is known to reduce breast cancer risk by ~50% in *BRCA1* and *BRCA2* mutation carriers (19), we chose to account for the remaining risk of cancer in women undergoing prophylactic oophorectomy by assessing it as an additional time-varying covariate rather than by censoring the follow-up of the women at the time they underwent this procedure. In addition, we did a sensitivity analysis to assess the potential for survival bias in our analyses by restricting the study to women more likely to have incident cases of breast cancer. Although no change in the significance of the results was observed following this approach, it is important to evaluate this possibility in any study, whether single site or multicenter, of individuals at significantly elevated risk of cancer.

This report represents the largest association study conducted to date in *BRCA1* and *BRCA2* carriers. It also is the first report from CIMBA, an international consortium established to provide sufficient statistical power to test candidate single nucleotide polymorphisms as modifiers of cancer risk in *BRCA1* and *BRCA2* mutation carriers and to refine breast cancer risk prediction in this population. The operating principles of CIMBA are as follows. (*a*) CIMBA is open to any group that can contribute genotype and phenotype information on at least 92 *BRCA1* and/or *BRCA2* mutation carriers. Groups with smaller collections of carriers are encouraged to participate through partnership with a larger group. (*b*) Phenotypic data obtained from risk factor questionnaires and/or medical records are uniformly coded and stored in a centralized CIMBA database. These data include year of birth, mutation description, ethnicity, country of residence, age at last follow-up, ages at breast and ovarian cancer diagnosis, age at bilateral prophylactic mastectomy, age at bilateral prophylactic oophorectomy, and status and age at menopause. (*c*) Panels of single nucleotide polymorphisms for genotyping are selected every 6 months at a CIMBA group meeting. (*d*) Only single nucleotide polymorphisms that show significant associations, either in the published literature or in data available to a member group, at $P < 0.01$, are considered. (*e*) Each investigator/group is free to participate or not in any round of genotyping. (*f*) Genotyping quality control standards must be followed (2% duplicates, call rates >95%, randomized arrangement of affected and unaffected carriers for genotyping). (*g*) Genotyping data from participating centers are pooled and analyzed as outlined in the CIMBA analysis plan. This study represents the first genetic modifier study conducted by CIMBA using these guidelines.

This study of 7,187 *BRCA1* and *BRCA2* carriers had 80% power to detect significant ($P < 0.05$) protective recessive effects with HRs of ≤0.82 for the F31I allele. We therefore conclude that the present study has a sufficient sample size to assess with reasonable confidence the involvement of the F31I allele in the modification of breast cancer risk among BRCA1 and BRCA2 mutation carries. It also shows the importance of large consortia, such as CIMBA, in evaluating the associations between genetic markers and cancer risk.

Appendix 1. Study Collaborators

MAGIC collaborators: Susan Neuhausen, University of California Irvine, Irvine, CA; Timothy Rebbeck, Susan Domchek, Katherine Nathanson, University of Pennsylvania School of Medicine, Philadelphia, PA; Theresa Wagner, Medical University of Vienna, Vienna, Austria; Judy Garber, Dana-Farber Cancer Institute, Boston, MA; Henry Lynch, Creighton University, Omaha, NE; Claudine Isaacs, Lombardi Cancer Center, Georgetown University, Washington, DC; Jeffrey Weitzel, City of Hope Cancer Center, Duarte, CA; Olufunmilayo Olopade, University of Chicago, Chicago, IL; Steven Narod, Centre for Research in Women's Health, Toronto, Ontario, Canada; Mary Daly and Andrew Godwin, Fox Chase Cancer Center, Philadelphia, PA; Gail Tomlinson, University of Texas Southwestern Medical Center at Dallas, Dallas, TX; Fergus Couch, Mayo Clinic, Rochester, MN.

GEMO study collaborators: Agnès Chompret, Brigitte Bressac-de-Paillerets, Véronique Byrde, Corinne Capoulade, Gilbert Lenoir, Institut Gustave Roussy, Villejuif, France; Yves-Jean Bignon, Nancy Uhrhammer, Centre Jean Perrin, Clermont-Ferrand, France; Marion Gauthier-Villars, Muriel Belotti, Antoine de Pauw, Dominique Stoppa-Lyonnet, Institut Curie, Paris, France; Laure Barjhoux, Mélanie Léone, Sophie Giraud, Olga Sinilnikova, Hospices Civils de Lyon/Centre Léon Bérard, Lyon, France; Christine Lasset, Valérie Bonadona, Centre Léon Bérard, Lyon, France; Agnès Hardouin, Pascaline Berthet, Centre François Baclesse, Caen, France; Hagay Sobol, Institut Paoli Calmettes, Marseille, France; Florence Coulet, Chrystelle Colas, Florent Soubrier, Hopital Pitié-Salpétrière, Paris, France; Isabelle Coupier, CHU de Arnaud-de-Villeneuve, Montpellier, France; Jean-Philippe Peyrat, Joëlle Fournier, Philippe Vennin, Claude Adenis, Centre Oscar Lambret, Lille, France; Catherine Nogues, Centre René Huguenin, St. Cloud, France; Rosette Lidereau, Institut National de la Sante et de la Recherche Medicale U735, Centre René Huguenin, St. Cloud, France; Danièle Muller, Jean-Pierre Fricker, Centre Paul Strauss, Strasbourg, France; Michel Longy, Institut Bergonié, Bordeaux, France; Christine Toulas, Rosine Guimbaud, Laurence Gladieff, Viviane Feillel, Institut Claudius

Regaud, Toulouse, France; Sylvie Mazoyer, Centre National de la Recherche Scientifique UMR5201, Lyon, France; Henry T. Lynch, Creighton University, Omaha, NE; Drakoulis Yannoukakos, National Center for Scientific Research Demokritos, Athens, Greece.

EMBRACE collaborators: Coordinating Centre, Cambridge: Douglas Easton, Antonis Antoniou, Susan Peock, Margaret Cook; North of Scotland Regional Genetics Service, Aberdeen: Neva Haites, Helen Gregory; Northern Ireland Regional Genetics Service, Belfast: Patrick J. Morrison; West Midlands Regional Clinical Genetics Service, Birmingham: Trevor Cole, Carole McKeown; South West Regional Genetics Service, Bristol: Alan Donaldson; East Anglian Regional Genetics Service, Cambridge: Joan Paterson; Medical Genetics Services for Wales, Cardiff: Jonathon Gray; St. James's Hospital and National Centre for Medical Genetics, Dublin: Peter Daly, David Barton; South East of Scotland Regional Genetics Service, Edinburgh: Mary Porteus, Michael Steel; Peninsula Clinical Genetics Service, Exeter: Carole Brewer, Julia Rankin; West of Scotland Regional Genetics Service, Glasgow: Rosemarie Davidson, Victoria Murday; South East Thames Regional Genetics Service, London: Louise Izatt, Gabriella Pichert; North West Thames Regional Genetics Service, Harrow: Huw Dorkins; Leicestershire Clinical Genetics Service, Leicester: Richard Trembath; Yorkshire Regional Genetics Service, Leeds: Tim Bishop, Carol Chu; Merseyside and Cheshire Clinical Genetics Service, Liverpool: Ian Ellis; Manchester Regional Genetics Service, Manchester: Gareth Evans, Fiona Lalloo, Andrew Shenton; North East Thames Regional Genetics Service, London: James Mackay, Anne Robinson; Nottingham Centre for Medical Genetics, Nottingham: Susan Ritchie, Sandy Raeburn; Northern Clinical Genetics Service, Newcastle: Fiona Douglas, John Burn; Oxford Regional Genetics Service, Oxford: Sarah Durell; Department of Cancer Genetics, Royal Marsden Hospital: Ros Eeles; North Trent Clinical Genetics Service, Sheffield: Jackie Cook, Oliver Quarrell; South West Thames Regional Genetics Service, London: Shirley Hodgson; and Wessex Clinical Genetics Service, Southampton: Diana Eccles, Anneke Lucassen.

GCHBOC study collaborators: Beatrix Versmold and Rita Schmutzler, Division of Molecular Gyneco-Oncology, University of Cologne, Cologne, Germany; Christoph Engel, Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany; Alfons Meindl, Department of Gynaecology and Obstetrics, Technical University, Munich, Germany; Christian Sutter, Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany; Jurgen Horst, Institute of Human Genetics, University of Muenster, Muenster, Germany; Dieter Schaefer, Institute of Human Genetics, University of Frankfurt, Frankfurt, Germany; Norbert Arnold, University of Schleswig-Holstein, Campus Kiel, Kiel, Germany; Wera Hofmann, Max Delbrück Center for Molecular Medicine, Berlin, Germany; Dieter Niederacher, University of Düsseldorf, Düsseldorf, Germany; Helmut Deissler, University of Ulm, Ulm, Germany; Karin Kast, University of Dresden, Dresden, Germany.

kConFab collaborators: Georgia Chenevix-Trench, Amanda Spurdle, http://www.kconfab.org/Organisation/Members.shtml

MOD-SQUAD collaborators: Michal Zikan, Petr Pohlreich, Zdenek Kleibl, Charles University, Prague, Czech Republic; Lenka Foretova, Machackova Eva, Lukesova Miroslava, Masaryk Memorial Cancer Institute, Brno, Czech Republic.

The LUMC collaborators: Peter Devilee, Maaike P.G. Vreeswijk, Hans F.A. Vasen, H. Meijers-Heijboer, and D. Halley.

HEBCS: Heli Nevanlinna, Johanna Tommiska, Kristiina Aittomaki, Carl Blomqvist, Kirsimari Aaltonen, Anmitta Tamminen, Helsinki University Central Hospital, Helsinki, Finland.

INHERIT BRCAs collaborators: Paul Bessette, Service de Gynécologie, Centre Hospitalier Universitaire de Sherbrooke, Fleurimont, Quebec, Canada; Peter Bridge, Molecular Diagnostic Laboratory, Alberta Children's Hospital, Calgary, Alberta, Canada; Jocelyne Chiquette and Louise Provencher, Clinique des Maladies du sein Deschênes-Fabia, Hôpital du saint-Sacrement, Quebec City, Quebec, Canada; Rachel Laframboise, Service de Médecine Génétique, CHUQ, Pavillon CHUL, Quebec City, Quebec, Canada; Jean Lépine, Centre Hospitalier Regional de Rimouski, Rimouski, Quebec, Canada; Bernard Lespérance and Roxane Pichette, Service d'hémato-oncologie, Hôpital du Sacré-Coeur, Montréal, Quebec, Canada; Marie Plante, Service de Gynécologie, CHUQ, L'Hôtel-Dieu de Québec, Quebec City, Quebec, Canada; and Patricia Voyer, Clinique des maladies du sein, Carrefour de Santé de Jonquière, Jonquière, Quebec, Canada.

MAYO collaborators: Fergus J. Couch, Noralane Lindor, Linda Wadum, Kiley Johnson, Jennifer Mentlick, Janet Olson, Mayo Clinic College of Medicine, Rochester, MN.

Acknowledgments

We thank Jennifer Scott for assistance with preparation of the manuscript; Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow-up Study for their contributions to this resource; and the many families who contribute to kConFab. The LUMC team thanks K. Kroeze-Jansema for technical assistance. The HEBCS team thanks Drs. Carl Blomqvist and Kirsimari Aaltonen, as well as Anitta Tamminen, M.Sc., for their kind help.

References

1. Ewart-Toland A, Briassouli P, de Koning JP, et al. Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. Nat Genet 2003;34:403–12.
2. Zhang D, Hirota T, Marumoto T, et al. Cre-loxP-controlled periodic Aurora-A overexpression induces mitotic abnormalities and hyperplasia in mammary glands of mouse models. Oncogene 2004;23:8720–30.
3. Egan KM, Newcomb PA, Ambrosone CB, et al. STK15 polymorphism and breast cancer risk in a population-based study. Carcinogenesis 2004;25:2149–53.
4. Sun T, Miao X, Wang J, et al. Functional Phe31Ile polymorphism in Aurora A and risk of breast carcinoma. Carcinogenesis 2004;25:2225–30.
5. Lo YL, Yu JC, Chen ST, et al. Breast cancer risk associated with genotypic polymorphism of the mitosis-regulating gene Aurora-A/STK15/BTAK. Int J Cancer 2005;115:276–83.
6. Ewart-Toland A, Dai Q, Gao YT, et al. Aurora-A/STK15 T+91A is a general low penetrance cancer susceptibility gene: a meta-analysis of multiple cancer types. Carcinogenesis 2005;26:1368–73.
7. Cox DG, Hankinson SE, Hunter DJ. Polymorphisms of the AURKA (STK15/Aurora kinase) gene and breast cancer risk (United States). Cancer Causes Control 2006;17:81–3.
8. Dai Q, Cai QY, Shu XO, et al. Synergistic effects of STK15 gene polymorphisms and endogenous estrogen exposure in the risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2004;13:2065–70.
9. Fletcher O, Johnson N, Palles C, et al. Inconsistent association between the STK15 F31I genetic polymorphism and breast cancer risk. J Natl Cancer Inst 2006;98:1014–8.
10. Xu X, Weaver Z, Linke SP, et al. Centrosome amplification and a defective G_2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells. Mol Cell 1999;3:389–95.
11. Deng CX. Roles of BRCA1 in centrosome duplication. Oncogene 2002;21:6222–7.
12. Wu K, Hinson SR, Ohashi A, et al. Functional evaluation and cancer risk assessment of BRCA2 unclassified variants. Cancer Res 2005;65:417–26.
13. Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro ANA, Tavtigian SV, Couch FJ. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet 2004;75:535–44.
14. Chenevix-Trench G, Healey S, Lakhani S, et al. Genetic and histopathologic evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown clinical significance. Cancer Res 2006;66:2019–27.
15. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: chose of the time-scale. Am J Epidemiol 1997;145:72–80.
16. Antoniou AC, Goldgar DE, Andrieu N, et al. A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes. Genet Epidemiol 2005;29:1–11.
17. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. J Am Stat Assoc 1989;84:1074–8.
18. Ouchi M, Fujiuchi N, Sasai K, et al. BRCA1 phosphorylation by Aurora-A in the regulation of G_2 to M transition. J Biol Chem 2004;279:19643–8.
19. Rebbeck TR, Lynch HT, Neuhausen SL, et al.; Prevention and Observation of Surgical End Points Study Group. Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations. N Engl J Med 2002;346:1616–22.

Commentary

An international initiative to identify genetic modifiers of cancer risk in *BRCA1* and *BRCA2* mutation carriers: the Consortium of Investigators of Modifiers of *BRCA1* and *BRCA2* (CIMBA)

Georgia Chenevix-Trench[1], Roger L Milne[2], Antonis C Antoniou[3], Fergus J Couch[4], Douglas F Easton[3] and David E Goldgar[5], on behalf of CIMBA

[1]Queensland Institute for Medical Research, Brisbane, Australia
[2]Human Cancer Genetics Program, Spanish National Cancer Centre (CNIO), Madrid, Spain
[3]CR-UK Genetic Epidemiology Unit, University of Cambridge, Cambridge, UK
[4]Mayo Clinic College of Medicine, Rochester, Minnesota, USA
[5]Department of Dermatology, University of Utah, Salt Lake City, Utah, USA

Corresponding author: Georgia Chenevix-Trench, georgiaT@qimr.edu.au

Published: 19 April 2007
This article is online at http://breast-cancer-research.com/content/9/2/104

Breast Cancer Research 2007, **9:**104 (doi:10.1186/bcr1670)

Abstract

BRCA1 and *BRCA2* mutations exhibit variable penetrance that is likely to be accounted for, in part, by other genetic factors among carriers. However, studies aimed at identifying these factors have been limited in size and statistical power, and have yet to identify any convincingly validated modifiers of the *BRCA1* and *BRCA2* phenotype. To generate sufficient statistical power to identify modifier genes, the Consortium of Investigators of Modifiers of *BRCA1* and *BRCA2* (CIMBA) has been established. CIMBA contains about 30 affiliated groups who together have collected DNA and clinical data from approximately 10,000 *BRCA1* and 5,000 *BRCA2* mutation carriers. Initial efforts by CIMBA to identify modifiers of breast cancer risk for *BRCA1* and *BRCA2* mutation carriers have focused on validation of common genetic variants previously associated with risk in smaller studies of carriers or unselected breast cancers. Future studies will involve replication of findings from pathway-based and genome-wide association studies in both unselected and familial breast cancer. The identification of genetic modifiers of breast cancer risk for *BRCA1* and *BRCA2* mutation carriers will lead to an improved understanding of breast cancer and may prove useful for the determination of individualized risk of cancer amongst carriers.

The search for genetic modifiers of *BRCA1* and *BRCA2*

Female carriers of deleterious *BRCA1* and *BRCA2* mutations are predisposed to high lifetime risks of breast and ovarian cancer. Initial estimates indicated that around 80% of carriers of mutations in *BRCA1* and *BRCA2* from multiple-case families would develop breast cancer by age 70 [1,2], and genetic counseling is usually carried out on the assumption that penetrance estimates apply to all women. However, a later pooled analysis from population-based studies estimated an average risk by age 70 in this context of 66% in *BRCA1* carriers and 45% in *BRCA2* carriers [3]. It has also been reported that cancer risks vary by the age at diagnosis and the type of cancer in the index case [3,4]. Such observations are consistent with the more plausible hypothesis that cancer risks in mutation carriers are modified by genetic factors or other risk factors that cluster in families. Segregation analysis has also demonstrated that models that allow for other genes to have a modifying effect on the breast cancer risks conferred by *BRCA1* and *BRCA2* mutations fit significantly better than models without a modifying component [5]. Further evidence for genetic modifiers arises from studies of risk factors that are themselves influenced by genetic factors. For example, mammographic density that has a strong genetic component [6] has been recently shown in one study to modify the breast cancer risks in *BRCA1* and *BRCA2* mutation carriers [7].

Although there has been considerable interest in finding genetic modifiers of cancer risk in *BRCA1* and *BRCA2* mutation carriers, the number of published studies is still fairly modest and has focused around genes involved in a limited number of pathways: detoxification of environmental carcinogens, DNA repair and steroidogenesis. Several studies have evaluated the CAG repeat length polymorphism in the androgen receptor (*AR*) gene as a modifier of breast cancer risk among mutation carriers. However, the data from different studies are contradictory and no firm conclusions can be drawn as to the magnitude of such an effect, if any [8-11]. Many studies have also evaluated a repeat length polymorphism in *AIB1* as a modifier of risk among *BRCA1* or

CIMBA = Consortium of Investigators of Modifiers of *BRCA1* and BRCA2; SNP = single nucleotide polymorphism.

BRCA2 mutation carriers. Although an effect of high numbers of repeats on cancer risk in carriers was first reported by Rebbeck and colleagues [12], three large subsequent studies failed to replicate this result [13-15]. *RAD51* currently provides the most convincing evidence for the existence of a modifier gene, at least for *BRCA2* mutation carriers. Levy-Lahad and colleagues [16] first reported that the -135G>C single nucleotide polymorphism (SNP) in the 5′ untranslated region of *RAD51* modified the breast cancer risk in *BRCA2* carriers and this finding has been substantiated by others [17,18]. The function of the -135G>C SNP in *RAD51* is not clear, but it could affect mRNA stability or translational efficiency.

Choosing candidate SNPs or genes to evaluate as modifiers of *BRCA1* and *BRCA2* suffers from the same problem faced by all candidate-based genetic association studies, namely the poor understanding of the relevant pathways and hence the small *a priori* likelihood that any of them are true modifiers [19]. These issues may be overcome in the future through the identification of candidate genomic regions associated with breast cancer risk by linkage analyses [20], or more plausibly by the identification of candidate SNPs by adequately powered genome-wide association studies [21]. In addition, the publication of convincingly validated SNPs associated with breast cancer in the general population [22] will provide some new candidates to test as modifiers of breast cancer risk among *BRCA1* or *BRCA2* mutation carriers. However, since SNPs associated with breast cancer in the general population may not act in the same way among *BRCA1* and *BRCA2* mutation carriers, pathway-based and perhaps genome-wide association studies in *BRCA1* and *BRCA2* carriers are also needed.

Consortium of Investigators of Modifiers of *BRCA1* and *BRCA2* (CIMBA)

A number of large studies and consortia have been established that aim to identify genetic modifiers of cancer risk in *BRCA1* and *BRCA2* mutation carriers, including Modifiers and Genetics in Cancer (MAGIC), Epidemiological study of BRCA1 and BRCA2 mutation carriers (EMBRACE), Genetic Modifiers of cancer risk in *BRCA1/2* mutation carriers (GEMO), the Kathleen Cuningham Consortium for Research into Familial Breast Cancer (kConFab), the German Consortium for Hereditary Breast and Ovarian Cancer (GCHBOC) and the Breast Cooperative Family Registry (Breast-CFR). However, with current sample sizes of less than 1,500 carriers, none of these groups have adequate power to identify genetic modifiers with confidence. To address this problem, a 'consortium of consortia', the Consortium of Investigators of Modifiers of *BRCA1* and *BRCA2* (CIMBA), was established in 2005 (see Additional file 1 for a list of current contributors). The operating principles of CIMBA are: CIMBA is open to any group that can contribute genotypic and basic phenotypic and epidemiological risk factor data from at least 100 female *BRCA1* and *BRCA2* mutation carriers with or without a cancer diagnosis - groups with smaller collections of carriers are encouraged to participate through partnership with a larger group; panels of SNPs for genotyping are selected at face-to-face meetings every six months; only SNPs that show significant associations (arbitrarily set at $p < 0.01$) with breast cancer risk in carriers, either in the published literature or in data from a member group, or are convincingly identified as associated with breast cancer in the general population, are considered; each group is free to participate, or not, in any round of genotyping; genotyping quality control standards must be followed (>2% duplicates, call rates >95%, no-template controls on every plate and randomized arrangement of affected and unaffected carriers for genotyping); all epidemiological risk factor data and genotyping data from carriers are submitted to the CIMBA data coordinating centre at the University of Cambridge; and genotyping data from participating centers are pooled for analysis. There are currently about 30 groups from North America, Europe and Australia who plan to contribute to some or all of the collaborative CIMBA projects, and collectively they have DNA and minimum required clinical and epidemiological data from more than 10,000 *BRCA1* and 5,000 *BRCA2* carriers.

Statistical considerations

Most association studies are case-control studies, in which genotype frequencies in a series of cases are compared with those in series of controls. The analysis of *BRCA1* and *BRCA2* modifiers is potentially more complex, because a high proportion of carriers become affected. Thus, modifiers would be expected to influence not just whether a carrier became affected but also the age at diagnosis. More powerful analyses can, therefore, be conducted by treating breast cancer as a survival (age at onset) rather than a simple binary endpoint. An additional problem, however, is introduced by the fact that mutation carriers are mainly ascertained through cancer genetics clinics. In these settings, the first tested individual in a family is usually someone diagnosed with cancer at a relatively young age. Such study designs tend, therefore, to lead to an over-sampling of affected individuals and standard analytical methods like Cox regression may lead to biased estimates of the risk ratios [5]. CIMBA aims to address this potential bias by using standard analytical methods, such as weighted Cox regression, or by analyzing the data within a retrospective likelihood framework [5]. In addition, analyses restricted to incident cases, defined as carriers diagnosed with cancer no more than five years prior to ascertainment, are applied to account in part for ascertainment and possible survival bias. One of the aims of CIMBA is also to further develop the statistical methodology used to analyze such data. Among *BRCA1* mutation carriers and at a threshold of $p < 0.0001$, CIMBA currently has a power of over 80% to detect polymorphisms with minor allele frequencies greater than 10% that confer risk ratios in excess of 1.2 (Table 1). The power is somewhat lower among the current sample of

Table 1

Simulated power (%) to detect a polymorphism with varying minor allele frequency and risk ratio, under a multiplicative model at a significance level 10^{-4}

Minor allele frequency	Relative hazard	Sample size: 5000	Sample size: 10,000
0.10	1.1	2	7
	1.2	33	80
	1.3	86	100
0.20	1.1	5	26
	1.2	74	100
	1.3	100	100
0.30	1.1	10	44
	1.2	89	100
	1.3	100	100

Simulations performed as in [5].

BRCA2 mutation carriers. However, it is still far greater than the power that be achieved by each study individually - at a minor allele frequency of 20% and risk ratio of 1.2, the corresponding power would be <5% for a sample size of approximately 1,000 carriers. Moreover, most of the participating CIMBA centers are actively recruiting carriers, and larger sample sizes are expected in the future.

Conclusions

The identification of convincingly validated modifiers of breast cancer risk for *BRCA1* and *BRCA2* mutation carriers will help to understand the biology of hereditary breast tumors and, in the case of *BRCA1*-mutation-associated risk modifiers, will also provide candidate low penetrance genes for 'sporadic' basal cell breast cancers because of their similarity to *BRCA1*-related breast tumors [23,24]. In the long term it might be possible to include information on genetic modifiers in risk prediction models, to give individualized advice to mutation carriers on individual breast cancer risks, and to have sufficient power to evaluate the risk of other cancers in *BRCA1* and *BRCA2* mutation carriers.

Additional file

The following Additional file is available online:

Additional file 1
Current contributors to CIMBA.
See http://breast-cancer-research.com/content/supplementary/bcr1670-s1.doc

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Statistical and data management support for CIMBA, and support for EMBRACE and A Antoniou, are provided by Cancer Research UK. The MAGIC Consortium is supported by US Public Health Service grants R01-CA102776, R01-CA08574 and P50-CA83638; EMBRACE by a Cancer Research-UK grant to DF Easton; kConFab by grants from the National Breast Cancer Foundation, the National Health and Medical Research Council (NHMRC, including grants 145684 and 288704) and by the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia; INHERIT by the Canadian Institutes of Health Research, and by the Canada Research Chair in Oncogenetics (J Simard); DNA-HEBON Study Netherlands by a grant from the Dutch Cancer Society, NKI 2007-3756; OCGN by Cancer Care Ontario, and OCGN and Susan Neuhausen and Saundra Buys through cooperative agreements with the Northern California Cancer Center, Cancer Care Ontario, The University of Melbourne, Fox Chase Cancer Center, Columbia University and the Huntsman Cancer Institute as part of the Breast Cancer Family Registry (Breast CFR) funded under RFA#CA-95-003; the Helsinki Breast Cancer Study by the Academy of Finland (110663), Finnish Cancer Society, Helsinki University Central Hospital Research Fund and the Sigrid Juselius Fund; the GEMO study by the Programme Hospitalier de Recherche Clinique AOR01082, by Programme Incitatif et Coopératif Génétique et Biologie de Cancer du Sein, Instutut Curie, and by the Association "Le cancer du sein, parlons-en!" Award; The GCHBOC by the German Cancer Aid (grant 107054) and the Center for Molecular Medicine Cologne (grant TV 93 to Rita Schmutzler).

G Chenevix-Trench and A Spurdle and supported by grants from the National Health and Medical Research Council of Australia; FJ Couch by grants from the Breast Cancer Research Foundation (BCRF), US Army Medical Research and Materiel Command (W81XWH-04-1-0588) and the Mayo Clinic Breast Cancer SPORE (P50-CA116201); K Nathanson by a grant from the Breast Cancer Research Foundation; C Szabo by the Susan G Komen Breast Cancer Foundation (BCTR0402923); M Zikan, P Pohlreich and Z Kleibel by the Research Project of the Ministry of Education, Youth and Sports of the Czech Republic (MSM0021620808); J Struewing in part by the Intramural Research Program of the NIH, Division of Cancer Epidemiology and Genetics and the Center for Cancer Research, National Cancer Institute, US Department of Health and Human Services.

References

1. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, *et al.*: **Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium.** *Am J Hum Genet* 1998, **62:**676-689.
2. Easton DF, Ford D, Bishop DT: **Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium.** *Am J Hum Genet* 1995, **56:**265-271.
3. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, *et al.*: **Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies.** *Am J Hum Genet* 2003, **72:**1117-1130.
4. Simchoni S, Friedman E, Kaufman B, Gershoni-Baruch R, Orr-Urtreger A, Kedar-Barnes I, Shiri-Sverdlov R, Dagan E, Tsabari S, Shohat M, *et al.*: **Familial clustering of site-specific cancer risks associated with BRCA1 and BRCA2 mutations in the Ashkenazi Jewish population.** *Proc Natl Acad Sci USA* 2006, **103:**3770-3774.
5. Antoniou AC, Goldgar DE, Andrieu N, Chang-Claude J, Brohet R, Rookus MA, Easton DF: **A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes.** *Genet Epidemiol* 2005, **29:**1-11.
6. Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, McCredie MR, Giles GG, Tritchler D, Chiarelli A, Yaffe MJ, *et al.*: **Heritability of mammographic density, a risk factor for breast cancer.** *N Engl J Med* 2002, **347:**886-894.
7. Mitchell G, Antoniou AC, Warren R, Peock S, Brown J, Davies R, Mattison J, Cook M, Warsi I, Evans DG, *et al.*: **Mammographic density and breast cancer risk in BRCA1 and BRCA2 mutation carriers.** *Cancer Res* 2006, **66:**1866-1872.

8. Dagan E, Friedman E, Paperna T, Carmi N, Gershoni-Baruch R: **Androgen receptor CAG repeat length in Jewish Israeli women who are BRCA1/2 mutation carriers: association with breast/ovarian cancer phenotype.** *Eur J Hum Genet* 2002, **10:**724-728.
9. Kadouri L, Easton DF, Edwards S, Hubert A, Kote-Jarai Z, Glaser B, Durocher F, Abeliovich D, Peretz T, Eeles RA: **CAG and GGC repeat polymorphisms in the androgen receptor gene and breast cancer susceptibility in BRCA1/2 carriers and non-carriers.** *Br J Cancer* 2001, **85:**36-40.
10. Rebbeck TR, Kantoff PW, Krithivas K, Neuhausen S, Blackwood MA, Godwin AK, Daly MB, Narod SA, Garber JE, Lynch HT, *et al.*: **Modification of BRCA1-associated breast cancer risk by the polymorphic androgen-receptor CAG repeat.** *Am J Hum Genet* 1999, **64:**1371-1377.
11. Spurdle AB, Antoniou AC, Duffy DL, Pandeya N, Kelemen L, Chen X, Peock S, Cook MR, Smith PL, Purdie DM, *et al.*: **The androgen receptor CAG repeat polymorphism and modification of breast cancer risk in BRCA1 and BRCA2 mutation carriers.** *Breast Cancer Res* 2005, **7:**R176-183.
12. Rebbeck TR, Wang Y, Kantoff PW, Krithivas K, Neuhausen SL, Godwin AK, Daly MB, Narod SA, Brunet JS, Vesprini D, *et al.*: **Modification of BRCA1- and BRCA2-associated breast cancer risk by AIB1 genotype and reproductive history.** *Cancer Res* 2001, **61:**5420-5424.
13. Hughes DJ, Ginolhac SM, Coupier I, Barjhoux L, Gaborieau V, Bressac-de-Paillerets B, Chompret A, Bignon YJ, Uhrhammer N, Lasset C, *et al.*: **Breast cancer risk in BRCA1 and BRCA2 mutation carriers and polyglutamine repeat length in the AIB1 gene.** *Int J Cancer* 2005, **117:**230-233.
14. Kadouri L, Kote-Jarai Z, Easton DF, Hubert A, Hamoudi R, Glaser B, Abeliovich D, Peretz T, Eeles RA: **Polyglutamine repeat length in the AIB1 gene modifies breast cancer susceptibility in BRCA1 carriers.** *Int J Cancer* 2004, **108:**399-403.
15. Spurdle AB, Antoniou AC, Kelemen L, Holland H, Peock S, Cook MR, Smith PL, Greene MH, Simard J, Plourde M, *et al.*: **The AIB1 polyglutamine repeat does not modify breast cancer risk in BRCA1 and BRCA2 mutation carriers.** *Cancer Epidemiol Biomarkers Prev* 2006, **15:**76-79.
16. Levy-Lahad E, Lahad A, Eisenberg S, Dagan E, Paperna T, Kasinetz L, Catane R, Kaufman B, Beller U, Renbaum P, *et al.*: **A single nucleotide polymorphism in the RAD51 gene modifies cancer risk in BRCA2 but not BRCA1 carriers.** *Proc Natl Acad Sci USA* 2001, **98:**3232-3236.
17. Kadouri L, Kote-Jarai Z, Hubert A, Durocher F, Abeliovich D, Glaser B, Hamburger T, Eeles RA, Peretz T: **A single-nucleotide polymorphism in the RAD51 gene modifies breast cancer risk in BRCA2 carriers, but not in BRCA1 carriers or noncarriers.** *Br J Cancer* 2004, **90:**2002-2005.
18. Wang WW, Spurdle AB, Kolachana P, Bove B, Modan B, Ebbers SM, Suthers G, Tucker MA, Kaufman DJ, Doody MM, *et al.*: **A single nucleotide polymorphism in the 5′ untranslated region of RAD51 and risk of cancer among BRCA1/2 mutation carriers.** *Cancer Epidemiol Biomarkers Prev* 2001, **10:**955-960.
19. Pharoah PD, Dunning AM, Ponder BA, Easton DF: **Association studies for finding cancer-susceptibility genetic variants.** *Nat Rev Cancer* 2004, **4:**850-860.
20. Nathanson KL, Shugart YY, Omaruddin R, Szabo C, Goldgar D, Rebbeck TR, Weber BL: **CGH-targeted linkage analysis reveals a possible BRCA1 modifier locus on chromosome 5q.** *Hum Mol Genet* 2002, **11:**1327-1332.
21. Hirschhorn JN, Daly MJ: **Genome-wide association studies for common diseases and complex traits.** *Nat Rev Genet* 2005, **6:** 95-108.
22. Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, Scollen S, Baynes C, Ponder BA, Chanock S, *et al.*: **A common coding variant in CASP8 is associated with breast cancer risk.** *Nat Genet* 2007, **39:**352-358.
23. Lakhani SR, Jacquemier J, Sloane JP, Gusterson BA, Anderson TJ, van de Vijver MJ, Farid LM, Venter D, Antoniou A, Storfer-Isser A, *et al.*: **Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations.** *J Natl Cancer Inst* 1998, **90:**1138-1145.
24. Lakhani SR, Reis-Filho JS, Fulford L, Penault-Llorca F, van der Vijver M, Parry S, Bishop T, Benitez J, Rivas C, Bignon YJ, *et al.*: **Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype.** *Clin Cancer Res* 2005, **11:**5175-5180.

ARTICLES

Genome-wide association study identifies novel breast cancer susceptibility loci

Douglas F. Easton[1], Karen A. Pooley[2], Alison M. Dunning[2], Paul D. P. Pharoah[2], Deborah Thompson[1], Dennis G. Ballinger[3], Jeffery P. Struewing[4], Jonathan Morrison[2], Helen Field[2], Robert Luben[5], Nicholas Wareham[5], Shahana Ahmed[2], Catherine S. Healey[2], Richard Bowman[6], the SEARCH collaborators[2]*, Kerstin B. Meyer[7], Christopher A. Haiman[8], Laurence K. Kolonel[9], Brian E. Henderson[8], Loic Le Marchand[9], Paul Brennan[10], Suleeporn Sangrajrang[11], Valerie Gaborieau[10], Fabrice Odefrey[10], Chen-Yang Shen[12], Pei-Ei Wu[12], Hui-Chun Wang[12], Diana Eccles[13], D. Gareth Evans[14], Julian Peto[15], Olivia Fletcher[16], Nichola Johnson[16], Sheila Seal[17], Michael R. Stratton[17,18], Nazneen Rahman[17], Georgia Chenevix-Trench[19], Stig E. Bojesen[20], Børge G. Nordestgaard[20], Christen K. Axelsson[21], Montserrat Garcia-Closas[22], Louise Brinton[22], Stephen Chanock[23], Jolanta Lissowska[24], Beata Peplonska[25], Heli Nevanlinna[26], Rainer Fagerholm[26], Hannaleena Eerola[26,27], Daehee Kang[28], Keun-Young Yoo[28,29], Dong-Young Noh[28], Sei-Hyun Ahn[30], David J. Hunter[31,32], Susan E. Hankinson[32], David G. Cox[31], Per Hall[33], Sara Wedren[33], Jianjun Liu[34], Yen-Ling Low[34], Natalia Bogdanova[35,36], Peter Schürmann[36], Thilo Dörk[36], Rob A. E. M. Tollenaar[37], Catharina E. Jacobi[38], Peter Devilee[39], Jan G. M. Klijn[40], Alice J. Sigurdson[41], Michele M. Doody[41], Bruce H. Alexander[42], Jinghui Zhang[4], Angela Cox[43], Ian W. Brock[43], Gordon MacPherson[43], Malcolm W. R. Reed[44], Fergus J. Couch[45], Ellen L. Goode[45], Janet E. Olson[45], Hanne Meijers-Heijboer[46,47], Ans van den Ouweland[47], André Uitterlinden[48], Fernando Rivadeneira[48], Roger L. Milne[49], Gloria Ribas[49], Anna Gonzalez-Neira[49], Javier Benitez[49], John L. Hopper[50], Margaret McCredie[51], Melissa Southey[50], Graham G. Giles[52], Chris Schroen[53], Christina Justenhoven[54], Hiltrud Brauch[54], Ute Hamann[55], Yon-Dschun Ko[56], Amanda B. Spurdle[19], Jonathan Beesley[19], Xiaoqing Chen[19], kConFab[57]*, AOCS Management Group[19,57]*, Arto Mannermaa[58,59], Veli-Matti Kosma[58,59], Vesa Kataja[58,60], Jaana Hartikainen[58,59], Nicholas E. Day[5], David R. Cox[3] & Bruce A. J. Ponder[2,7]

Breast cancer exhibits familial aggregation, consistent with variation in genetic susceptibility to the disease. Known susceptibility genes account for less than 25% of the familial risk of breast cancer, and the residual genetic variance is likely to be due to variants conferring more moderate risks. To identify further susceptibility alleles, we conducted a two-stage genome-wide association study in 4,398 breast cancer cases and 4,316 controls, followed by a third stage in which 30 single nucleotide polymorphisms (SNPs) were tested for confirmation in 21,860 cases and 22,578 controls from 22 studies. We used 227,876 SNPs that were estimated to correlate with 77% of known common SNPs in Europeans at $r^2 > 0.5$. SNPs in five novel independent loci exhibited strong and consistent evidence of association with breast cancer ($P < 10^{-7}$). Four of these contain plausible causative genes (*FGFR2*, *TNRC9*, *MAP3K1* and *LSP1*). At the second stage, 1,792 SNPs were significant at the $P < 0.05$ level compared with an estimated 1,343 that would be expected by chance, indicating that many additional common susceptibility alleles may be identifiable by this approach.

Breast cancer is about twice as common in the first-degree relatives of women with the disease as in the general population, consistent with variation in genetic susceptibility to the disease[1]. In the 1990s, two major susceptibility genes for breast cancer, *BRCA1* and *BRCA2*, were identified[2,3]. Inherited mutations in these genes lead to a high risk of breast and other cancers[4]. However, the majority of multiple case breast cancer families do not segregate mutations in these genes. Subsequent genetic linkage studies have failed to identify further major breast cancer genes[5]. These observations have led to the proposal that breast cancer susceptibility is largely 'polygenic': that is, susceptibility is conferred by a large number of loci, each with a small effect on breast cancer risk[6]. This model is consistent with the observed patterns of familial aggregation of breast cancer[7]. However, progress in identifying the relevant loci has been slow. As linkage studies lack power to detect alleles with moderate effects on risk, large case-control association studies are required. Such studies have identified variants in the DNA repair genes *CHEK2*, *ATM*, *BRIP1* and *PALB2* that confer an approximately twofold risk of breast cancer, but these variants are rare in the population[8–14]. A recent study has shown that a common coding variant in *CASP8* is associated with a moderate reduction in breast cancer risk[15]. After accounting for all the known breast cancer loci, more than 75% of the familial risk of the disease remains unexplained[16].

Recent technological advances have provided platforms that allow hundreds of thousands of SNPs to be analysed in association studies, thus providing a basis for identifying moderate risk alleles without

Affiliations of the above authors are given at the end of the paper.
*Lists of consortia participants and affiliations appear after author affiliations.

prior knowledge of position or function. It has been estimated that there are 7 million common SNPs in the human genome (with minor allele frequency, m.a.f., >5%)[17]. However, because recombination tends to occur at distinct 'hot-spots', neighbouring polymorphisms are often strongly correlated (in 'linkage disequilibrium', LD) with each other. The majority of common genetic variants can therefore be evaluated for association using a few hundred thousand SNPs as tags for all the other variants[18]. We aimed to identify further breast cancer susceptibility loci in a three-stage association study[19]. In the first stage, we used a panel of 266,722 SNPs, selected to tag known common variants across the entire genome[18]. These SNPs were genotyped in 408 breast cancer cases and 400 controls from the UK; data were analysed for 390 cases and 364 controls genotyped for ≥80% of the SNPs. The cases were selected to have a strong family history of breast cancer, equivalent to at least two affected female first-degree relatives, because such cases are more likely to carry susceptibility alleles[20]. Initally, we analysed 227,876 SNPs (85%) with genotypes on at least 80% of the subjects. We estimate that these SNPs are correlated with 58% of common SNPs in the HapMap CEPH/CEU (Utah residents with ancestry from northern and western Europe) samples at $r^2 > 0.8$, and 77% at $r^2 > 0.5$ (mean $r^2 = 0.75$; see Supplementary Fig. 1) (http://www.hapmap.org/)[21]. As expected, coverage was strongly related to m.a.f.: 70% of SNPs with m.a.f. > 10% were tagged at $r^2 > 0.8$, compared with 23% of SNPs with m.a.f. 5–10%. The main analyses were restricted to 205,586 SNPs that had a call rate of 90% and whose genotype distributions did not differ from Hardy–Weinberg equilibrium in controls (at $P < 10^{-5}$).

For the second stage we selected 12,711 SNPs, approximately 5% of those typed in stage 1, on the basis of the significance of the difference in genotype frequency between cases and controls. These SNPs were then genotyped in a further 3,990 invasive breast cancer cases and 3,916 controls from the SEARCH study, using a custom-designed oligonucleotide array. In the main analyses, we considered 10,405 SNPs with call rate of >95% that did not deviate from Hardy–Weinberg equilibrium in controls.

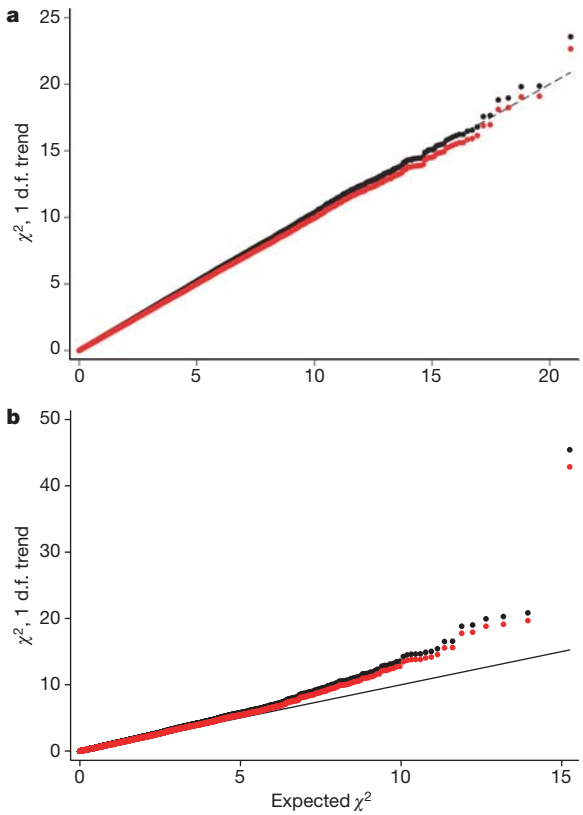

Figure 1 | Quantile–quantile plots for the test statistics (Cochran-Armitage 1 d.f. χ^2 trend tests) for stages 1 and 2. a, Stage 1; **b**, stage 2. Black dots are the uncorrected test statistics. Red dots are the statistics corrected by the genomic control method ($\lambda = 1.03$ for stage 1, $\lambda = 1.06$ for stage 2). Under the null hypothesis of no association at any locus, the points would be expected to follow the black line.

Comparison of the observed and expected distribution of test statistics showed some evidence for an inflation of the test statistics in both stage 1 (inflation factor $\lambda = 1.03$, 95% confidence interval (CI) 1.02–1.04) and stage 2 ($\lambda = 1.06$, 95% CI 1.04–1.12), based on the 90% least significant SNPs (Fig. 1). Possible explanations for this inflation include population stratification, cryptic relatedness among subjects, and differential genotype calling between cases and controls. There was evidence for an excess of low call rate SNPs among the most significant SNPs ($P < 0.01$) in stage 1, but not in stage 2, suggesting that some of this effect is a genotyping artefact (Supplementary Table 1). However, the inflation was still present among SNPs with call rate >99% in both cases and controls, possibly reflecting population substructure. We computed 1 degree of freedom (d.f.) association tests for each SNP, combining stages 1 and 2. After adjustment for this inflation by the genomic control method[22], we observed more associations than would have been expected by chance at $P < 0.05$ (Table 1). One SNP (dbSNP rs2981582) was significant at the $P < 10^{-7}$ level that has been proposed as appropriate for genome-wide studies[23].

In the third stage, to establish whether any SNPs were definitely associated with risk, we tested 30 of the most significant SNPs in 22 additional case-control studies, comprising 21,860 cases of invasive breast cancer, 988 cases of carcinoma *in situ* (CIS) and 22,578 controls (Supplementary Table 2). Six SNPs showed associations in stage 3 that were significant at $P \leq 10^{-5}$ with effects in the same direction as in stages 1 and 2 (Table 2, Supplementary Table 3, and Fig. 2). All these SNPs reached a combined significance level of $P < 10^{-7}$ (ranging from 2×10^{-76} to 3×10^{-9}). Of these six SNPs, five were within genes or LD blocks containing genes. SNP rs2981582 lies in intron 2 of *FGFR2* (also known as *CEK3*), which encodes the fibroblast growth factor receptor 2. SNPs rs12443621 and rs8051542 are both located in an LD block containing the 5′ end of *TNRC9* (also known as *TOX3*), a gene of uncertain function containing a tri-nucleotide repeat motif, as well as the hypothetical gene, *LOC643714*. SNP rs889312 lies in an LD block of approximately 280 kb that contains *MAP3K1* (also known as *MEKK*), which encodes the signalling protein mitogen-activated protein kinase kinase kinase 1, in addition to two other genes: *MGC33648* and *MIER3*. SNP rs3817198 lies in intron 10 of *LSP1* (also known as *WP43*), encoding lymphocyte-specific protein 1, an F-actin bundling cytoskeletal protein expressed in haematopoietic and endothelial cells. A further SNP, rs2107425, located just 110 kilobases (kb) from rs3817198, was also identified (overall $P = 0.00002$). rs2107425 is within the *H19* gene, an imprinted maternally expressed untranslated messenger RNA closely involved in regulation of the insulin growth factor gene, *IGF2*. In stage 3, however, rs2107425 was only weakly significant after adjustment for rs3817198 by logistic regression ($P = 0.06$). This suggests that the association with breast cancer risk may be driven by variants in *LSP1* rather than in *H19*. The sixth SNP reaching a combined $P < 10^{-7}$ was rs13281615, which lies on 8q. It is correlated with SNPs in a 110 kb LD block that contains no known

Table 1 | Number of significant associations after stage 2

Level of significance	Observed	Observed adjusted*	Expected	Ratio
0.01–0.05	1,239	1,162	934.3	1.24
0.001–0.01	574	517	347.6	1.49
0.0001–0.001	112	88	53.3	1.65
0.00001–0.0001	16	12	7.0	1.71
<0.00001	15	13	0.96	13.5
All $P < 0.05$	1,956	1,792	1,343.2	1.33

Observed numbers of SNPs associated with breast cancer after stage 2, by level of significance, before and after adjustment for population stratification, and expected numbers under the null hypothesis of no association.
* Adjusted for inflation of the test statistic by the genomic control method.

Table 2 | Summary of results for eleven SNPs selected for stage 3 that showed evidence of an association with breast cancer

rs Number	Gene	Position*	m.a.f.†	Per allele OR (95% CI)	HetOR (95% CI)	HomOR (95% CI)	*P*-trend		
							Stages 1 and 2	Stage3	Combined
rs2981582	*FGFR2*	10q 123342307	0.38 *(0.30)*	1.26 (1.23–1.30)	1.23 (1.18–1.28)	1.63 (1.53–1.72)	4×10^{-16}	5×10^{-62}	2×10^{-76}
rs12443621	*TNRC9/ LOC643714*	16q 51105538	0.46 *(0.60)*	1.11 (1.08–1.14)	1.14 (1.09–1.20)	1.23 (1.17–1.30)	10^{-7}	9×10^{-14}	2×10^{-19}
rs8051542	*TNRC9/ LOC643714*	16q 51091668	0.44 *(0.20)*	1.09 (1.06–1.13)	1.10 (1.05–1.16)	1.19 (1.12–1.27)	4×10^{-6}	4×10^{-8}	10^{-12}
rs889312	*MAP3K1*	5q 56067641	0.28 *(0.54)*	1.13 (1.10–1.16)	1.13 (1.09–1.18)	1.27 (1.19–1.36)	4×10^{-6}	3×10^{-15}	7×10^{-20}
rs3817198	*LSP1*	11p 1865582	0.30 *(0.14)*	1.07 (1.04–1.11)	1.06 (1.02–1.11)	1.17 (1.08–1.25)	8×10^{-6}	10^{-5}	3×10^{-9}
rs2107425	*H19*	11p 1977651	0.31 *(0.44)*	0.96 (0.93–0.99)	0.94 (0.90–0.98)	0.95 (0.89–1.01)	7×10^{-6}	0.01	2×10^{-5}
rs13281615		8q 128424800	0.40 *(0.56)*	1.08 (1.05–1.11)	1.06 (1.01–1.11)	1.18 (1.10–1.25)	2×10^{-7}	6×10^{-7}	5×10^{-12}
rs981782		5p 45321475	0.47 *(0.37)*	0.96 (0.93–0.99)	0.96 (0.92–1.01)	0.92 (0.87–0.97)	8×10^{-5}	0.003	9×10^{-6}
rs30099		5q 52454339	0.08 *(0.39)*	1.05 (1.01–1.10)	1.06 (1.00–1.11)	1.09 (0.96–1.24)	0.003	0.02	0.001
rs4666451		2p 19150424	0.41 *(0.04)*	0.97 (0.94–1.00)	0.98 (0.93–1.02)	0.93 (0.87–0.99)	5×10^{-6}	0.04	6×10^{-5}
rs3803662‡	*TNRC9/ LOC643714*	16q 51143842	0.25 *(0.60)*	1.20 (1.16–1.24)	1.23 (1.18–1.29)	1.39 (1.26–1.45)	3×10^{-12}	10^{-26}	10^{-36}

OR, odds ratio; HetOR, odds ratio in heterozygotes; HomOR, odds ratio in rare homozygotes (relative to common homozygotes); CI, confidence interval.
* Build 36.2 position.
† Minor allele frequency in SEARCH (UK) study. Combined allele frequency from three Asian studies in italics.
‡ rs3803662 was not part of the initial tag SNP set but identified as a result of fine-scale mapping of the *TNRC9/LOC643714* locus and typed in the stage 2 and stage 3 sets (but not the stage 1 set).

genes. The basis of this association therefore remains obscure. This SNP is approximately 130 kb proximal to rs1447295, 60 kb proximal to rs6983267 and 230 kb distal to rs16901979, recently shown to be associated with prostate cancer[24–26].

In addition to the seven SNPs described above, there was evidence of association among the remaining 23 SNPs (global $P = 0.001$ in stage 3). In particular, three SNPs showed some evidence of association in stage 3 ($P < 0.05$, in each case in the same direction as in stages 1 and 2; Table 2). SNPs rs981782 and rs30099 both lie in the centromeric region of chromosome 5. rs4666451 lies on 2p, a region for which some evidence of linkage to breast cancer in families has been reported[5]. The 20 other SNPs showed no evidence of association in stage 3 (global $P = 0.11$), suggesting that most of these associations from stages 1 and 2 were false positives.

FGFR2

The most significantly associated SNP, rs2981582, lies within a 25 kb LD block almost entirely within intron 2 of *FGFR2*. We found no evidence of association with SNPs elsewhere in the gene (Fig. 3a). In an attempt to identify a causal variant, we first identified the 19 common variants (m.a.f. > 0.05) in this block from HapMap CEU data. These were tagged ($r^2 > 0.8$) by 7 SNPs including rs2981582. The additional tag SNPs were genotyped in the SEARCH study cases and controls. Multiple logistic regression analysis of these variants found no additional evidence for association after adjusting for rs2981582. Haplotype analysis of these 7 SNPs indicated that multiple haplotypes carrying the minor (*a*) allele of rs2981582 were associated with an increased risk of breast cancer, implying that the association was being driven by rs2981582 itself or a variant strongly correlated with it (Supplementary Table 4).

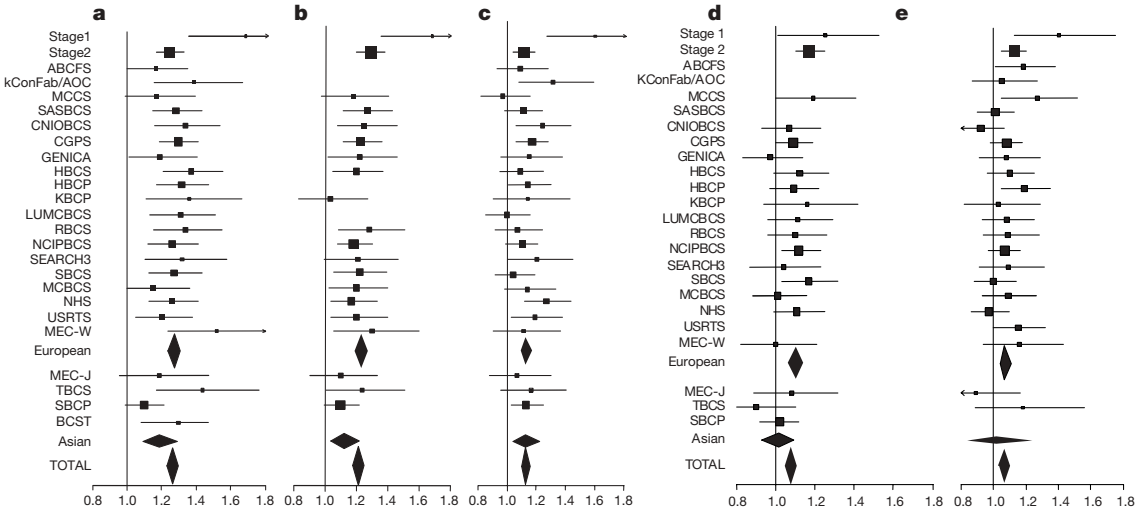

Figure 2 | Forest plots of the per-allele odds ratios for each of the five SNPs reaching genome-wide significance. a, rs2981582; **b**, rs3803662; **c**, rs889312; **d**, rs13281615; and **e**, rs3817198. The *x*-axis gives the per-allele odds ratio. Each row represents one study (see Supplementary Table 2), with summary odds ratios for all European and all Asian studies, and all studies combined. The area of the square for each study is proportional to the inverse of the variance of the estimate. Horizontal lines represent 95% confidence intervals. Diamonds represent the summary odds ratios, with 95% confidence intervals, based on the stage 3 studies only.

Resequencing of this region in 45 subjects of European origin identified 29 variants that were strongly correlated with rs2981582 ($r^2 > 0.6$) (http://cgwb.nci.nih.gov; Fig. 3b and Supplementary Tables 5–8). A subset of 14 variants tagged 27 of these in European ($r^2 > 0.95$) and Asian (Korean) samples ($r^2 > 0.86$). Two variants could not be genotyped reliably. This new tagging set was then genotyped in SEARCH and 3 studies from Asian populations; the Asian studies were included because the LD is weaker, providing greater power to resolve the causal variant (Fig. 3b, left panel). The strongest association was found with rs7895676. On the assumption that there is a single disease-causing allele, we calculated a likelihood for each variant. 21 SNPs (including rs2981582) had a likelihood ratio of <1/100 relative to rs7895676, indicating that none of these are likely to be the causal variant (Supplementary Table 8). Six variants were too strongly correlated for their individual effects to be separated using a genetic epidemiological approach. Functional assays will be required to determine which is causally related to breast cancer risk.

Intron 2 of *FGFR2* shows a high degree of conservation in mammals, and contains several putative transcription-factor binding sites (http://genomequebec.mcgill.ca/PReMod)[27], some of which lie in close proximity to the relevant SNPs. We therefore speculate that the association with breast cancer risk is mediated through regulation of *FGFR2* expression. Of possible relevance is that only three of these variants (rs10736303, rs2981578 and rs35054928) are within sequences conserved across all placental mammals (Fig. 3c and Supplementary Table 8). Of these, the disease associated allele of rs10736303 generates a putative oestrogen receptor (ER) binding site. rs35054928 lies immediately adjacent to a perfect POU domain protein octamer (Oct) binding site. However, multiple splice variants have been reported in *FGFR2*, and differential splicing might provide an alternative mechanism for the association. FGFR2 is a receptor tyrosine kinase that is amplified and overexpressed in 5–10% of breast tumours[28–30]. Somatic missense mutations of *FGFR2* that are likely to be implicated in cancer development have also been demonstrated in primary tumours and cell lines of multiple tumour types (http://www.sanger.ac.uk/genetics/CGP/cosmic/)[30,31].

TNRC9/LOC643714 locus

As two SNPs in the *TNRC9/LOC643714* locus, rs12443621 and rs8051542, both showed convincing evidence of association, we further evaluated this region by genotyping, in the SEARCH set, an additional 19 SNPs tagging 101 common variants within the entire *TNRC9* and *LOC643714* genes, based on the HapMap CEU data. SNPs tagging the coding region of *TNRC9* showed no evidence of association. The strongest association was observed with rs3803662, a synonymous coding SNP of *LOC643714* that lies 8 kb upstream of *TNRC9*. This SNP was therefore genotyped in the stage 3 set (Table 2). Logistic regression analysis indicated that rs3803662 exhibited a stronger association with disease than other SNPs, and the associations with other SNPs were non-significant after adjustment for rs3803662. These results suggest

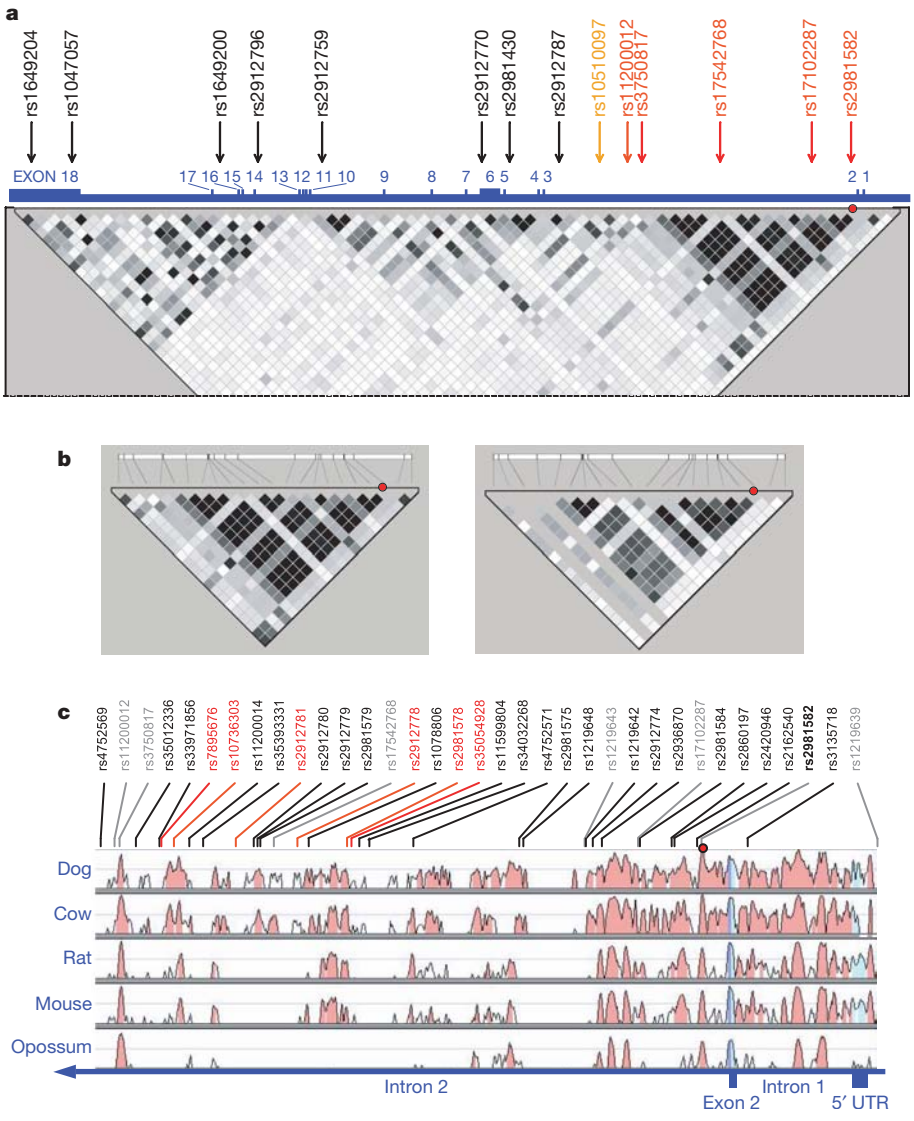

Figure 3 | The *FGFR2* locus. a, Map of the whole *FGFR2* gene, viewed relative to common SNPs on HapMap. The gene is 126 kb long and in reverse 3′–5′ orientation on chromosome 10. Exon positions are illustrated with respect to the 67 SNPs with m.a.f. > 5% in HapMap CEU (therefore the map is not to physical scale). Numbered SNPs are those tested in the genome-wide study. SNPs in black were not significant in stage 1. Those in red were significant at $P < 0.0001$ after stage 2. rs10510097 (orange) was significant in stage 1, but failed quality control in stage 2 owing to deviation from Hardy–Weinberg equilibrium. Squares indicate pairwise r^2 on a greyscale (black = 1, white = 0). Red circle indicates rs2981582. **b**, Resequenced 32 kb region, shown relative to SNPs in CEU with m.a.f. > 5%, showing pairwise LD for SNPs in HapMap CEU (left panel) and JPT/CHB (right panel). Red circle indicates rs2981582, shown in bold black. **c**, Sequence conservation of 32 kb region in five species, relative to human sequence (http://pipeline.lbl.gov/methods.shtml)[35]. Red circle indicates rs2981582. SNPs in grey are those used in the initial tagging of known common HapMap SNPs within the block. SNPs in black are correlated with rs2981582 with $r^2 > 0.6$ in European samples. Six SNPs in red were those consistent with being the causative variant on the basis of the genetic data (not excluded at odds of 100:1 relative to the SNP with the strongest association, rs7895676).

that the causal variant is closely correlated with rs3803662. Four SNPs in the HapMap CEU data (rs17271951, rs1362548, rs3095604 and rs4784227) that span *LOC643714* and the 5′ regulatory regions of *TNRC9* are strongly correlated with rs3803662, and it therefore remains unclear in which gene the causative variant lies. *TNRC9* contains a putative HMG (high mobility group) box motif, suggesting that it might act as a transcription factor.

Pattern of risks

We assessed in more detail, in the stage 3 data, the pattern of the risks associated with the five independent SNPs that reached an overall $P < 10^{-7}$: rs2981582 (*FGFR2*), rs3803662 (*TNRC9/LOC643714*), rs889312 (*MAP3K1*), rs13281615 (8q) and rs3817198 (*LSP1*). For each of these five SNPs, the minor allele in Europeans was associated with an increased risk of breast cancer in a dose-dependent manner, with a higher risk of breast cancer in homozygous than in heterozygous carriers. Simple dominant and recessive models could be rejected for each SNP (all $P = 0.02$ or less). There was a marked difference in allele frequencies between populations, with the risk-associated alleles of rs8051542, rs889312 and rs13281615 being the major allele in Asian populations. The per allele odds ratio associated with rs2981582 was significantly smaller, though still elevated, in the Asian versus European populations ($P = 0.04$ for difference in odds ratio). This difference is consistent with the hypothesis that rs2981582 is not the functional variant at the *FGFR2* locus, and was not seen for SNPs exhibiting stronger evidence in the fine-scale mapping. No other evidence for heterogeneity in the per-allele odds ratio among studies was observed (Fig. 2).

Three of the SNPs (rs2981582, rs3803662 and rs889312) also showed evidence of association with breast CIS (Supplementary Table 9). For rs2981582 and rs3803662, the estimated odds ratios were greater for a diagnosis of breast cancer before age 40 years, but the trends by age were not statistically significant (Supplementary Table 10). There was evidence of an association with family history of breast cancer for three SNPs: for rs2981582 ($P = 0.02$), rs3803662 ($P = 0.03$) and rs13281615 ($P = 0.05$), the susceptibility allele was commoner in women with a first-degree relative with the disease than in those without (Supplementary Table 11). rs2981582 was also associated with bilaterality ($P = 0.02$). The associations with family history and bilaterality are to be expected for susceptibility loci, and are similar to previous observations for alleles in *CHEK2* and *ATM* (refs 10, 12, 14).

Discussion

This study has identified five novel breast cancer susceptibility loci, and demonstrated conclusively that some of the variation in breast cancer risk is due to common alleles. None of the loci we identified had been previously reported in association studies. Most previously identified breast cancer susceptibility genes are involved in DNA repair, and many association studies in breast cancer have concentrated on genes in DNA repair and sex hormone synthesis and metabolism pathways. None of the associations reported here appear to relate to genes in these pathways. It is notable that three of the five loci contain genes related to control of cell growth or to cell signalling, but only one (*FGFR2*) had a clear prior relevance to breast cancer. These results should, therefore, open up new avenues for basic research.

Our results emphasize the critical importance of study size in genetic association studies. It is notable that none of the confirmed associations reached genome-wide significance after stage 1 and only one reached this level after stage 2. As most common cancers have similar familial relative risks to breast cancer, it is likely that similarly large studies will be required to identify common alleles for other cancers. The fine-scale mapping of the *FGFR2* locus demonstrates that, even with a clear association, identification of the causative variant can be extremely problematic. However, the use of studies from multiple populations with different patterns of LD can substantially reduce the number of variants that need to be subjected to functional analysis.

As these susceptibility alleles are very common, a high proportion of the general population are carriers of at-risk genotypes. For example, approximately 14% of the UK population and 19% of UK breast cancer cases are homozygous for the rare allele at rs2981582. On the other hand, the increased risks associated with these alleles are relatively small—on the basis of UK population rates, the estimated breast cancer risk by age 70 years for rare homozygotes at rs2981582 is 10.5%, compared to 6.7% in heterozygotes and 5.5% in common homozygotes. At this stage, it is unlikely that these SNPs will be appropriate for predictive genetic testing, either alone or in combination with each other. However, as further susceptibility alleles are identified, a combination of such alleles together with other breast cancer risk factors may become sufficiently predictive to be important clinically.

On the basis of the relative risk estimates from stage 3, and assuming that the five most significant loci interact multiplicatively on disease risk, these loci explain an estimated 3.6% of the excess familial risk of breast cancer. On the basis of our staged design and the estimated distribution of linkage disequilibrium between the typed SNPs and those in HapMap, we estimate that the power to identify the five most significant associations at $P < 10^{-7}$ (rs2981582, rs3803662, rs889312, rs13281615 and rs3817198) was 93%, 71%, 25%, 3% and 1% respectively. These estimates are uncertain, notably because the true coverage of HapMap SNPs is unknown. Nevertheless, these calculations indicate that the power to detect the two strongest associations was high, and suggest that there are likely to be few other common variants with a similar effect on variation in breast cancer risk to rs2981582. In contrast, the low power to detect rs13281615 and rs3817198 suggests that these variants may represent a much larger class of loci, each explaining of the order of 0.1% of the familial risk of breast cancer. An example of such a locus is provided by *CASP8* D302H, which showed strong evidence of association in a previous large study[15]. This SNP was tested in stage 1, but the association was missed because it did not reach the threshold for testing in stage 2. The excess of associations after stage 2 is also consistent with the existence of many such loci. In addition, because the coverage for SNPs with m.a.f. $< 10\%$ was low, many low frequency alleles may have been missed. The detection of further susceptibility loci will require genome-wide studies with more complete coverage and using larger numbers of cases and controls, together with the combination of results across multiple studies. The present study demonstrates that common susceptibility loci can be reliably identified, and that they may together explain an appreciable fraction of the genetic variance in breast cancer risk.

METHODS SUMMARY

Cases for stage 1 were identified through clinical genetics centres in the UK and a national study of bilateral breast cancer. Cases in stage 2 were drawn from a population-based study of breast cancer (SEARCH)[32]. Controls for stages 2 and 3 were drawn from EPIC-Norfolk, a population-based study of diet and cancer[33].

Cases and controls for stage 3 were identified through case-control studies in Europe, North America, South-East Asia and Australia participating in the Breast Cancer Association Consortium (Supplementary Table 2)[34].

Genotyping for stages 1 and 2 was conducted using high-density oligonucleotide microarrays. For the main analyses, we excluded samples called on ≤80% of SNPs in either stage. We also excluded SNPs that achieved a call rate of ≤90% in stage 1 and ≤95% in stage 2, and SNPs whose frequency deviated from Hardy–Weinberg equilibrium in controls at $P < 0.00001$. Genotyping for stage 3, and for the fine-scale mapping of the *FGFR2* locus, was conducted using either a 5′ nuclease assay (Taqman, Applied Biosystems) or MALDI-TOF mass spectrometry using the Sequenom iPLEX system. For each centre, we excluded any sample called on ≤80% of SNPs, and any SNP with a call rate of ≤95% or a deviation from Hardy–Weinberg equilibrium in controls at $P < 0.00001$. Tests of association were 1 d.f. Cochran-Armitage tests, stratified for stage, centre and ethnic group (European or Asian). Odds ratios for each SNP were estimated using stratified logistic regression, using the stage 3 data only.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 February; accepted 30 April 2007.
Published online 27 May 2007; corrected 28 June 2007 (details online).

1. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological

studies including 58 209 women with breast cancer and 101 986 women without the disease. *Lancet* **358,** 1389–1399 (2001).
2. Miki, Y. *et al.* A strong candidate for the breast and ovarian-cancer susceptibility gene BRCA1. *Science* **266,** 66–71 (1994).
3. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* **378,** 789–792 (1995).
4. Antoniou, A. *et al.* Average risks of breast and ovarian cancer associated with mutations in *BRCA1* or *BRCA2* detected in case series unselected for family history: A combined analysis of 22 studies. *Am. J. Hum. Genet.* **72,** 1117–1130 (2003).
5. Smith, P. *et al.* A genome wide linkage search for breast cancer susceptibility genes. *Genes Chromosom. Cancer* **45,** 646–655 (2006).
6. Pharoah, P. D. P. *et al.* Polygenic susceptibility to breast cancer and implications for prevention. *Nature Genet.* **31,** 33–36 (2002).
7. Antoniou, A. C., Pharoah, P. D. P., Smith, P. & Easton, D. F. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br. J. Cancer* **91,** 1580–1590 (2004).
8. Rahman, N. *et al. PALB2*, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature Genet.* **39,** 165–167 (2007).
9. Thompson, D. *et al.* Cancer risks and mortality in heterozygous ATM mutation carriers. *J. Natl Cancer Inst.* **97,** 813–822 (2005).
10. Meijers-Heijboer, H. *et al.* Low-penetrance susceptibility to breast cancer due to *CHEK2**1100delC in noncarriers of *BRCA1* or *BRCA2* mutations. *Nature Genet.* **31,** 55–59 (2002).
11. Erkko, H. *et al.* A recurrent mutation in *PALB2* in Finnish cancer families. *Nature* **446,** 316–319 (2007).
12. Renwick, A. *et al. ATM* mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature Genet.* **38,** 873–875 (2006).
13. Seal, S. *et al.* Truncating mutations in the Fanconi anemia J gene *BRIP1* are low-penetrance breast cancer susceptibility alleles. *Nature Genet.* **38,** 1239–1241 (2006).
14. The CHEK2 Breast Cancer Case-Control Consortium. CHEK2*1100delC and susceptibility to breast cancer: A collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from ten studies. *Am. J. Hum. Genet.* **74,** 1175–1182 (2004).
15. Cox, A. *et al.* A common coding variant in *CASP8* is associated with breast cancer risk. *Nature Genetics* 39, 352–358 (2007); corrigendum **39,** 688 (2007).
16. Easton, D. F. How many more breast cancer predisposition genes are there? *Breast Cancer Res.* **1,** 1–4 (1999).
17. Kruglyak, L. & Nickerson, D. A. Variation is the spice of life. *Nature Genet.* **27,** 234–236 (2001).
18. Hinds, D. A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* **307,** 1072–1079 (2005).
19. Satagopan, J. M., Verbel, D. A., Venkatraman, E. S., Offit, K. E. & Begg, C. B. Two-stage designs for gene-disease association studies. *Biometrics* **58,** 163–170 (2002).
20. Antoniou, A. C. & Easton, D. F. Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet. Epidemiol.* **25,** 190–202 (2003).
21. Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
22. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).
23. Thomas, D. C., Haile, R. W. & Duggan, D. Recent developments in genomewide association scans: A workshop summary and review. *Am. J. Hum. Genet.* **77,** 337–345 (2005).
24. Amundadottir, L. T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nature Genet.* **38,** 652–658 (2006).
25. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature Genet.* **39,** 645–649 (2007).
26. Gudmundsson, J. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nature Genet.* **39,** 631–637 (2007).
27. Ferretti, V. *et al.* PReMod: a database of genome-wide mammalian cis-regulatory module predictions. *Nucleic Acids Res.* **35,** D122–D126 (2007).
28. Moffa, A. B., Tannheimer, S. L. & Ethier, S. P. Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells. *Mol. Cancer Res.* **2,** 643–652 (2004).
29. Adnane, J. *et al.* Bek and Flg, 2 receptors to members of the Fgf family, are amplified in subsets of human breast cancers. *Oncogene* **6,** 659–663 (1991).
30. Jang, J. H., Shin, K. H. & Park, J. G. Mutations in fibroblast growth factor receptor 2 and fibroblast growth factor receptor 3 genes associated with human gastric and colorectal cancers. *Cancer Res.* **61,** 3541–3543 (2001).
31. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158 (2007).
32. Lesueur, F. *et al.* Allelic association of the human homologue of the mouse modifier *Ptprj* with breast cancer. *Hum. Mol. Genet.* **14,** 2349–2356 (2005).
33. Day, N. *et al.* EPIC-Norfolk: Study design and characteristics of the cohort. *Br. J. Cancer* **80,** 95–103 (1999).
34. Breast Cancer Association Consortium. Commonly studied SNPs and breast cancer: Negative results from 12,000 – 32,000 cases and controls from the Breast Cancer Association Consortium. *J. Natl Cancer Inst.* **98,** 1382–1396 (2006).
35. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res.* **30,** 38–41 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank the women who took part in this research, and all the funders and support staff who made this study possible. The principal funding for this study was provided by Cancer Research UK. Detailed acknowledgements are provided in Supplementary Information.

Author Contributions D.F.E., A.M.D., P.D.P.P., D.R.C. and B.A.J.P. designed the study and obtained financial support. D.G.B. and D.R.C. directed the genotyping of stages 1 and 2. D.F.E. and D.T. conducted the statistical analysis. K.A.P. and A.M.D. coordinated the genotyping for stage 3 and the fine-scale mapping of the *FGFR2* and *TNRC9* loci. J.P.S. and J.Z. performed resequencing and analysis of the *FGFR2* locus. K.A.P., S.A., C.S.H., R.B., C.A.H., L.K.K., B.E.H., L.L.M., P.B., S.S., V.G., F.O., C-Y. S., P-E.W. and H-C.W. conducted genotyping for the fine-scale mapping. R.L., J.M., H.F. and K.B.M. provided bioinformatics support. D.E., D.G.E., J.P., O.F., N.J., S.S., M.R.S. and N.R. coordinated the studies used in stage 1. N.W. and N.E.D. coordinated the EPIC study used in stages 1 and 2. The remaining authors coordinated the studies in stage 3 and undertook genotyping in those studies. D.F.E. drafted the manuscript, with substantial contributions from K.A.P., A.M.D., P.D.P.P. and B.A.J.P. All authors contributed to the final paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.F.E. (d.easton@srl.cam.ac.uk).

Author affiliations: [1]CR-UK Genetic Epidemiology Unit, Department of Public Health and Primary Care and, [2]Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK. [3]Perlegen Sciences, Inc., 2021 Stierlin Court, Mountain View, California 94043, USA. [4]Laboratory of Population Genetics, US National Cancer Institute, Bethesda, Maryland 20892, USA. [5]EPIC, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. [6]MRC Dunn Clinical Nutrition Centre, Cambridge CB2 0XY, UK. [7]Cancer Research UK Cambridge Research Institute, Cambridge CB2 0RE, UK. [8]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. [9]Epidemiology Program, Cancer Research Center of Hawaii, University of Hawaii, Honolulu, Hawaii 96813, USA. [10]International Agency for Research on Cancer, 150 Cours Albert Thomas, Lyon 69008, France. [11]National Cancer Institute, Bangkok 10400, Thailand. [12]Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan. [13]Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton SO16 5YA, UK. [14]Regional Genetic Service, St Mary's Hospital, Manchester M13 0JH, UK. [15]London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK, and Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. [16]Breakthrough Breast Cancer Research Centre, London SW3 6JB, UK. [17]Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. [18]Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [19]Queensland Institute of Medical Research, Brisbane, Queensland 4006, Australia. [20]Departments of Clinical Biochemistry and [21]Breast Surgery, Herlev and Bispebjerg University Hospitals, University of Copenhagen, DK-2730 Herlev, Denmark. [22]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland 20852, USA. [23]Advanced Technology Center, National Cancer Institute, Gaithersburg, Maryland 20877, USA. [24]Cancer Center and M. Sklodowska-Curie Institute of Oncology, Warsaw 02781, Poland. [25]Nofer Institute of Occupational Medicine, Lodz 90950, Poland. [26]Departments of Obstetrics and Gynecology, and [27]Department of Oncology, Helsinki University Central Hospital, Helsinki 00029, Finland. [28]Seoul National University College of Medicine, Seoul 151-742, Korea. [29]National Cancer Center, Goyang 411-769, Korea. [30]Ulsan University College of Medicine, Ulsan 680-749, Korea. [31]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, 677 Huntington Ave., Boston, Massachusetts 02115, USA. [32]Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Ave., Boston, Massachusetts 02115, USA. [33]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm SE-171 77, Sweden. [34]Population Genetics, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Republic of Singapore. [35]Department of Radiation Oncology and [36]Department of Gynecology and Obstetrics, Hannover Medical School, D-30625 Hannover, Germany. [37]Department of Surgery and [38]Department of Medical Decision Making and [39]Departments of Human Genetics and Pathology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands. [40]Family Cancer Clinic, Department of Medical Oncology, Erasmus MC-Daniel den Hoed Cancer Center, Groene Hilledijk 301, 3075 EA Rotterdam, the Netherlands. [41]Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, Maryland 20892, USA. [42]Environmental Health Sciences, University of Minnesota, Minneapolis, Minnesota 55455, USA. [43]Institute for Cancer Studies and [44]Academic Unit of Surgical Oncology, Sheffield University Medical School, Sheffield S10 2RX, UK. [45]Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. [46]VU University Medical Center, 1007 MB Amsterdam, the Netherlands. [47]Department of Clinical Genetics and [48]Internal Medicine, Erasmus University, Rotterdam NL-3015-GE, the Netherlands. [49]Spanish National Cancer Centre (CNIO), Madrid E-28029, Spain. [50]Centre for Molecular, Environmental, Genetic and Analytical Epidemiology, University of Melbourne, Carlton, Victoria 3053, Australia. [51]Department of Preventive and Social Medicine, University of Otago, Dunedin 9001, New Zealand. [52]Cancer Epidemiology Centre, Cancer Council Victoria, Carlton, Victoria 3053, Australia. [53]Genetic Epidemiology

Laboratory, Department of Pathology, University of Melbourne, Parkville, Victoria 3052, Australia. [54]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharamcology, 70376 Stuttgart and University of Tuebingen, 72074 Tuebingen, Germany. [55]Deutsches Krebsforschungszentrum, Heidelberg 69120, Germany. [56]Evangelische Kliniken Bonn gGmbH Johanniter Krankenhaus, 53113 Bonn, Germany. [57]Peter MacCallum Cancer Centre, Melbourne, Victoria 3002, Australia. [58]Insitute of Clinical Medicine, Pathology and Forensic Medicine, University of Kuopio, Kuopio FIN-70210, Finland. [59]Departments of Oncology and Pathology, University Hospital of Kuopio, Kuopio FIN-70211, Finland. [60]Department of Oncology, Vaasa Central Hospital, Vaasa 65130, Finland.

The SEARCH collaborators Craig Luccarini[1], Don Conroy[1], Mitul Shah[1], Hannah Munday[1], Clare Jordan[1], Barbara Perkins[1], Judy West[1], Karen Redman[1] & Kristy Driver[1]. **kConFab** Morteza Aghmesheh[2], David Amor[3], Lesley Andrews[4], Yoland Antill[5], Jane Armes[6], Shane Armitage[7], Leanne Arnold[7], Rosemary Balleine[8], Glenn Begley[9], John Beilby[10], Ian Bennett[11], Barbara Bennett[4], Geoffrey Berry[12], Anneke Blackburn[13], Meagan Brennan[14], Melissa Brown[15], Michael Buckley[16], Jo Burke[17], Phyllis Butow[18], Keith Byron[19], David Callen[20], Ian Campbell[21], Georgia Chenevix-Trench[22], Christine Clarke[23], Alison Colley[24], Dick Cotton[25], Jisheng Cui[26], Bronwyn Culling[27], Margaret Cummings[28], Sarah-Jane Dawson[5], Joanne Dixon[29], Alexander Dobrovic[30], Tracy Dudding[31], Ted Edkins[32], Maurice Eisenbruch[33], Gelareh Farshid[34], Susan Fawcett[35], Michael Field[36], Frank Firgaira[37], Jean Fleming[38], John Forbes[39], Michael Friedlander[40], Clara Gaff[41], Mac Gardner[41], Mike Gattas[42], Peter George[43], Graham Giles[44], Grantley Gill[45], Jack Goldblatt[46], Sian Greening[47], Scott Grist[37], Eric Haan[48], Marion Harris[49], Stewart Hart[50], Nick Hayward[22], John Hopper[51], Evelyn Humphrey[17], Mark Jenkins[52], Alison Jones[7], Rick Kefford[53], Judy Kirk[54], James Kollias[55], Sergey Kovalenko[56], Sunil Lakhani[57], Jennifer Leary[54], Jacqueline Lim[58], Geoff Lindeman[59], Lara Lipton[60], Liz Lobb[61], Mariette Maclurcan[62], Graham Mann[23], Deborah Marsh[63], Margaret McCredie[64], Michael McKay[49], Sue Anne McLachlan[65], Bettina Meiser[4], Roger Milne[26], Gillian Mitchell[49], Beth Newman[66], Imelda O'Loughlin[67], Richard Osborne[51], Lester Peters[68], Kelly Phillips[5], Melanie Price[62], Jeanne Reeve[69], Tony Reeve[70], Robert Richards[71], Gina Rinehart[72], Bridget Robinson[73], Barney Rudzki[74], Elizabeth Salisbury[75], Joe Sambrook[21], Christobel Saunders[76], Clare Scott[5], Elizabeth Scott[77], Rodney Scott[31], Ram Seshadri[37], Andrew Shelling[78], Melissa Southey[26], Amanda Spurdle[22], Graeme Suthers[48], Donna Taylor[79], Christopher Tennant[58], Heather Thorne[21], Sharron Townshend[46], Kathy Tucker[4], Janet Tyler[4], Deon Venter[80], Jane Visvader[81], Ian Walpole[46], Robin Ward[82], Paul Waring[30], Bev Warner[83], Graham Warren[67], Elizabeth Watson[67], Rachael Williams[84], Judy Wilson[85], Ingrid Winship[69] & Mary Ann Young[49]. **AOCS Management Group** David Bowtell[86], Adele Green[22], Anna deFazio[87], Georgia Chenevix-Trench[22], Dorota Gertig[51] & Penny Webb[22].

Consortia affiliations: [1]Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK. [2]Oncology Research Centre, Prince of Wales Hospital, Randwick, New South Wales 2031, Australia. [3]Genetic Health Services Victoria, Royal Children's Hospital, Melbourne, Victoria 3050, Australia. [4]Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick, New South Wales 2031, Australia. [5]Department of Haematology and Medical Oncology, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002, Australia. [6]Anatomical Pathology, Royal Women's Hospital, Carlton, Victoria 3053, Australia. [7]Molecular Genetics Laboratory, Royal Brisbane and Women's Hospital, Herston, Queensland 4029, Australia. [8]Departments of Translational and Medical Oncology, Westmead Hospital, Westmead, New South Wales 2145, Australia. [9]Cancer Biology Laboratory, TVW Institute for Child Health Research, Subiaco, Western Australia 6008, Australia. [10]Pathology Centre, Queen Elizabeth Medical Centre, Nedlands, Western Australia 6009, Australia. [11]Silverton Place, 101 Wickham Terrace, Brisbane, Queensland 4000, Australia. [12]Department of Public Health and Community Medicine, University of Sydney, Sydney, New South Wales 2006, Australia. [13]John Curtin School of Medical Research, Australian National University, PO Box 334, Canberra, Australian Capital Territory 2601, Australia. [14]NSW Breast Cancer Institute, PO Box 143, Westmead, New South Wales 2145, Australia. [15]Department of Biochemistry, University of Queensland, St. Lucia, Queensland 4072, USA. [16]Molecular and Cytogenetics Unit, Prince of Wales Hospital, Randwick, New South Wales 2031, Australia. [17]Royal Hobart Hospital, GPO Box 1061L, Hobart, Tasmania 7001, Australia. [18]Medical Psychology Unit, Royal Prince Alfred Hospital, Camperdown, New South Wales 2204, Australia. [19]Australian Genome Research Facility, Walter & Eliza Hall Medical Research Institute, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia. [20]Dame Roma Mitchell Cancer Research Laboratories, University of Adelaide/ Hanson Institute, PO Box 14, Rundle Mall, South Australia 5000, Australia. [21]Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. [22]Queensland Institute of Medical Research, Herston, Queensland 4006, Australia. [23]Westmead Institute for Cancer Research, University of Sydney, Westmead Hospital ,Westmead, New South Wales 2145, Australia. [24]Department of Clinical Genetics, Liverpool Health Service, PO Box 103, Liverpool, New South Wales 2170, Australia. [25]Mutation Research Centre, St Vincent's Hospital, Victoria Parade, Fitzroy, Victoria 3065, Australia. [26]Centre for Genetic Epidemiology, The University of Melbourne, Level 2 723 Swanston Street, Carlton, Victoria 3053, Australia. [27]Molecular and Clinical Genetics, Level 1 Building 65, Royal Prince Alfred Hospital, Camperdown, New South Wales 2050, Australia. [28]Department of Pathology, University of Queensland Medical School, Herston, New South Wales 4006, Australia. [29]Central Regional Genetic Services, Wellington Hospital, Private bag 7902, Wellington South 6039, New Zealand. [30]Molecular Department of Pathology, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. [31]Hunter Genetics, Hunter Area Health Service, Waratah, New South Wales 2310, Australia. [32]Clinical Chemistry, Princess Margret Hospital for Children, Box D184, Perth, Western Australia 6001, Australia. [33]Department of Multicultural Health, University of Sydney, New South Wales 2052, Australia. [34]Tissue Pathology, Institute of Medical & Veterinary Science, Adelaide, South Australia 5000, Australia. [35]Family Cancer Clinic, Monash Medical Centre, Clayton, Victoria 3168, Australia. [36]Faculty of Medicine, Royal North Shore Hospital, Vindin House, St Leonards, New South Wales 2065, Australia. [37]Department of Haematology, Flinders Medical Centre, Bedford Park, South Australia 5042, Australia. [38]Eskitis Institute of Cell & Molecular Therapies, School of Biomolecular and Biomedical Sciences, Griffith University, Nathan, Queensland 4111, Australia. [39]Surgical Oncology, University of Newcastle, Newcastle Mater Hospital, Waratah, New South Wales 2298, Australia. [40]Department of Medical Oncology, Prince of Wales Hospital, Randwick, New South Wales 2031, Australia. [41]Victorian Clinical Genetics Service, Royal Melbourne Hospital, Parkville, Victoria 3052, Australia. [42]Queensland Clinical Genetic Service, Royal Children's Hospital, Bramston Terrace, Herston, Queensland 4020, Australia. [43]Clinical Biochemistry Unit, Canterbury Health Labs, PO Box 151, Christchurch 8140, New Zealand. [44]Cancer Epidemiology Centre, The Cancer Council Victoria, 1 Rathdowne Street, Carlton, Victoria 3053, Australia. [45]Department of Surgery, Royal Adelaide Hospital, Adelaide, South Australia 5000, Australia. [46]Genetic Services of WA, King Edward Memorial Hospital, 374 Bagot Road, Subiaco, Western Australia 6008, Australia. [47]Wollongong Hereditary Cancer Clinic, Wollongong Public Hospital, Private Mail Bag 8808, South Coast Mail Centre, New South Wales 2521, Australia. [48]Department of Medical Genetics, Women's and Children's Hospital, North Adelaide, South Australia 5006, Australia. [49]Familial Cancer Clinic, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. [50]Breast and Ovarian Cancer Genetics, Monash Medical Centre, 871 Centre Road, Bentleigh East, Victoria 3165, Australia. [51]Centre for Molecular Environmental, Genetic & Analytic Epidemiology, University of Melbourne, Melbourne, Victoria 3010, Australia. [52]School of Population Health, The University of Melbourne, 723 Swanston Street, Carlton, Victoria 3053, Australia. [53]Medical Oncology, Westmead Hospital, Westmead, New South Wales 2145, Australia. [54]Familial Cancer Service, Department of Medicine, Westmead Hospital, Westmead, New South Wales 2145, Australia. [55]Breast Endocrine and Surgical Unit, Royal Adelaide Hospital, North Terrace, South Australia 5000, Australia. [56]Molecular Pathology Department, Southern Cross Pathology, Monash Medical Centre, Clayton, Victoria 3168, Australia. [57]Molecular and Cellular Pathology, The University of Queensland, Herston, Queensland 4006, Australia. [58]Department of Psychological Medicine, Royal North Shore Hospital, St Leonards, New South Wales 2065, Australia. [59]Breast Cancer Laboratory, Walter and Eliza Hall Institute, PO Royal Melbourne Hospital, Parkville, Victoria 3050, Australia. [60]Medical Oncology and Clinical Haematology Unit, Western Hospital, Footscray, Victoria 3011, Australia. [61]WA Centre for Cancer, Edith Cowan University, Churchlands, Western Australia 6018, Australia. [62]Department of Psychological Medicine, University of Sydney, New South Wales 2006, Australia. [63]Kolling Institute of Medical Research, Royal North Shore Hospital, St Leonards, New South Wales 2065, Australia. [64]Cancer Epidemiology Research Unit, NSW Cancer Council, 153 Dowling Street, Woolloomooloo, New South Wales 2011, Australia. [65]Department of Oncology, St Vincent's Hospital, 41 Victoria Parade, Fitzroy, Victoria 3065, Australia. [66]School of Public Health, Queensland University of Technology, Victoria Park, Kelvin Grove, Queensland 4059, Australia. [67]St Vincent's Breast Clinic, PO Box 4751, Toowoomba, Queensland 4350, Australia. [68]Radiation Oncology, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. [69]Genetic Services, Auckland Hospital, Private Bag 92024, Auckland 1142, New Zealand. [70]Cancer Genetics Laboratory, University of Otago, PO Box 56, Dunedin 9054, New Zealand. [71]Department of Cytogenetics and Molecular Genetics, Women and Children's Hospital, Adelaide, South Australia 5006, Australia. [72]Hancock Family Breast Cancer Foundation, PO Locked Bag 2, West Perth, Western Australia 6005, Australia. [73]Oncology Service, Christchurch Hospital, Private Bag 4710, Christchurch 8140, New Zealand. [74]Molecular Pathology Institute of Medical and Veterinary Science, Frome Road, Adelaide, South Australia 5000, Australia. [75]Section of Cytology, Institute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, New South Wales 2145, Australia. [76]School of Surgery and Pathology, QE11 Medical Centre, M block 2nd Floor, Nedlands, Western Australia 6907, Australia. [77]South View Clinic, Suite 13, Level 3 South Street, Kogarah, New South Wales 2217, Australia. [78]Department of Obstetrics and Gynaecology, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. [79]Department of Radiology, Royal Perth Hospital, Box X2213, Perth 6011, Western Australia, Australia. [80]Murdoch Institute, Royal Children's Hospital, Parkville, Victoria 3050, Australia. [81]Molecular Genetics of Cancer Division, Walter & Eliza Hall Medical Research Institute, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia. [82]Department of Medical Oncology, St Vincents Hospital, Darlinghurst, New South Wales 2010, Australia. [83]Cabrini Hospital, 183 Wattletree Road, Malvern, Victoria 3144, Australia. [84]Family Cancer Clinic, St Vincent's Hospital, Darlinghurst, New South Wales 2010, Australia. [85]Medical Psychology Research Unit, Royal North Shore Hospital, St Leonards, New South Wales 2065, Australia. [86]Cancer Genomics & Biochemistry Laboratory, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. [87]Obstetrics & Gynaecology, Westmead Hospital, University of Sydney, New South Wales 2006, Australia.

METHODS

Subjects. Cases in stage 1 were identified through clinical genetics centres in Cambridge ($n = 91$), Manchester (96) and Southampton (136), and a national study of bilateral breast cancer (85). Cases were women diagnosed with invasive breast cancer under the age of 60 years who had a family history score of at least 2, where the score was computed as the total number of first-degree relatives plus half the number of second-degree relatives affected with breast cancer. The score for women with bilateral breast cancer was increased by 1, so that women were eligible if they were diagnosed with bilateral breast cancer and had one affected first-degree relative. Cases known to carry a *BRCA1* or *BRCA2* mutation were excluded. Controls were selected from the EPIC-Norfolk study, a population-based cohort study of diet and cancer based in Norfolk, East Anglia, UK[33]. Controls were chosen to be women aged over 50 years and free of cancer at the time of entry. Genotyping was attempted on 408 cases, plus 32 duplicate case samples, and 400 controls. For the analysis in Table 1, 54 samples with genotype call rates <80% were excluded, so the final analyses were based on 390 cases and 364 controls. The minimum genotype call rate for the remaining samples was 89%. The overall genotype discordance rate between duplicate samples in stage 1 was 0.01%.

For stage 2, invasive breast cancer cases were drawn from SEARCH, a population-based study of cancer in East Anglia[32]. Controls were women selected from the EPIC-Norfolk study, as previously described[33]. Eighty-eight subjects who were also genotyped in stage 1, and 35 controls who subsequently developed breast cancer and were also in the case series, were excluded from the analysis, leaving 3,990 breast cancer cases and 3,916 controls, plus five duplicates. The overall rate of discordance of genotypes between duplicate samples in stage 2 was 0.008%.

Twenty-one additional studies were included in stage 3 (see Supplementary Table 2). These studies participated through the Breast Cancer Association Consortium, an ongoing collaboration among investigators conducting case-control association studies in breast cancer[15,33]. All studies provided information on disease status (invasive breast cancer, carcinoma *in situ* or control), age at diagnosis/observation, ethnic group, first-degree family history of breast cancer and bilaterality of breast cancer. One further study (Breast Cancer Study of Taiwan) was included in the fine-scale mapping of the *FGFR2* locus.

Genotyping. For stage 1, genotyping was performed on 200 ng DNA that was first subjected to whole genome amplification using Multiple Displacement Amplification (MDA)[36]. Samples were then genotyped for a set of 266,732 SNPs using high-density oligonucleotide, photolithographic microarrays at Perlegen Sciences. For stage 2, genotyping was performed using 2.5 μg genomic DNA. These samples were genotyped for a set of 13,023 SNPs selected on the basis of the stage 1 results, using a custom designed oligonucleotide array. For both stages, each SNP was interrogated by 24 25-mer oligonucleotide probes synthesized by photolithography on a glass substrate. The 24 features comprise 4 sets of 6 features interrogating the neighbourhoods of SNP reference and alternative alleles on forward and reference strands. Each allele and strand is represented by five offsets: −2, −1, 0, 1 and 2 indicating the position of the SNP within the 25-mer, with zero being at the thirteenth base. At offset 0 a quartet was tiled, which included the perfect match to reference and alternative SNP alleles, and the two remaining nucleotides as mismatch probes. When possible, the mismatch features were selected as a purine nucleotide substitution for a purine perfect match nucleotide and a pyrimidine nucleotide substitution for a pyrimidine perfect match nucleotide. Thus, each strand and allele tiling consisted of 6 features comprising five perfect match probes and one mismatch.

Individual genotypes were determined by clustering all SNP scans in the two-dimensional space defined by reference and alternative trimmed mean intensities, corrected for background. Allele frequencies were approximated using the intensities collected from the high-density oligonucleotide arrays. An SNP's allele frequency, p, was estimated as the ratio of the relative amount of the DNA with reference allele to the total amount of DNA. The $\hat{p}$ value was computed from the trimmed mean intensities of perfect match features, after subtracting a measure of background computed from trimmed means of intensities of mismatch features. The trimmed mean disregarded the highest and the lowest intensity from the five perfect match intensities before computing the arithmetic mean. For the mismatch features, the trimmed mean is the individual intensity of the specified mismatch feature.

The genotype clustering procedure was an iterative algorithm developed as a combination of K-means and constrained multiple linear regressions. The K-means at each step re-evaluated the cluster membership representing distinct diploid genotypes. The multiple linear regressions minimized the variance in $\hat{p}$ within each cluster while optimizing the regression lines' common intersect. The common intersect defined a measure of common background that was used to adjust the allele frequencies for the next step of K-means. The K-means and multiple linear regression steps were iterated until the cluster membership and background estimates converged. The best number of clusters was selected by maximizing the total likelihood over the possible cluster counts of 1, 2 and 3 (representing the combinations of the three possible diploid genotypes). The total likelihood was composed of data likelihood and model likelihood. The data likelihood was determined using a normal mixture model for the distribution of $\hat{p}$ around the cluster means. The model likelihood was calculated using a prior distribution of expected cluster positions, resulting in optimal $\hat{p}$ positions of 0.8 for the homozygous reference cluster, 0.5 for the heterozygous cluster and 0.2 for the homozygous alternative cluster.

A genotyping quality metric was compiled for each genotype from 15 input metrics that described the quality of the SNP and the genotype. The genotyping quality metric correlated with a probability of having a discordant call between the Perlegen platform and outside genotyping platforms (that is, non-Perlegen HapMap project genotypes). A system of 10 bootstrap aggregated regression trees was trained using an independent data set of concordance data between Perlegen genotypes and HapMap project genotypes. The trained predictor was then used to predict the genotyping quality for each of the genotypes in this data set. Genotypes with quality scores of less than 7 were discarded. Data were analysed for 227,876 SNPs in stage 1 and 12,026 (of 13,023 selected) in stage 2, for which the call rate was >80%.

The 12,711 SNPs for stage 2 were primarily selected on the basis of a 1 d.f. Cochran-Armitage trend test (11,809, all with $P < 0.052$). We also included 826 SNPs with $P < 0.01$ testing for the difference in frequency of either homozygote between cases and controls (that is, assuming either a dominant or recessive model) and 76 SNPs that achieved $P < 0.01$ on a Cochran-Armitage test, weighting individuals by their family history score as above.

For the main analyses, we discarded SNPs with a call rate <90% in stage 1 and 95% in stage 2, and SNPs with a deviation from Hardy–Weinberg equilibrium significant at $P < 0.00001$ in either stage, leaving 205,586 SNPs in stage 1 and 10,621 SNPs in stage 2.

The 30 SNPs included in the stage 3 analyses were initially selected on the basis of a combined analysis of stage 1 and stage 2. We included all SNPs achieving a combined $P < 0.00002$ (based on either the Cochran-Armitage or 2 d.f. test, see below). Following re-evaluation of the stage 2 genotyping by 5′ nuclease assay (Taqman, Applied Biosystems) using the ABI PRISM 7900HT (Applied Biosystems), and exclusion of some samples, 16 of these SNPs were significant at $P < 0.00002$ and 24 at $P < 0.0002$ (Supplementary Table 3). One additional SNP, rs3803662, was added as a result of fine-scale mapping of the *TNRC9*/*LOC643714* locus.

The 31 stage 3 SNPs were genotyped in 22 studies (including cases and controls from SEARCH not used in stage 2, together with 21 other studies). For 18 of the studies, genotyping was performed by 5′ nuclease assay (Taqman) using the ABI PRISM 7900HT or 7500 Sequence Detection Systems according to manufacturer's instructions. Primers and probes were supplied directly by Applied Biosystems (http://www.appliedbiosystems.com/) as Assays-by-Design. All assays were carried out in 384-well or 96-well format, with each plate including negative controls (with no DNA). Duplicate genotypes were provided for at least 2% of samples in each study. For three studies, SNPs were genotyped using matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) for the determination of allele-specific primer extension products using Sequenom's MassARRAY system and iPLEX technology. The design of oligonucleotides was carried out according to the guidelines of Sequenom and performed using MassARRAY Assay Design software (version 1.0). Multiplex PCR amplification of amplicons containing SNPs of interest was performed using Qiagen HotStart Taq Polymerase on a Perkin Elmer GeneAmp 2400 thermal cycler (MJ Research) with 5 ng genomic DNA. Primer extension reactions were carried out according to manufacturer's instructions for iPLEX chemistry. Assay data were analysed using Sequenom TYPER software (version 3.0). One study used both the Taqman and MALDI-TOF MS approaches. The SNPs genotyped in stage 3 were also regenotyped in the stage 2 samples using Taqman; these genotype calls were used in the overall analyses (Table 2, Supplementary Table 3, and Fig. 2).

We eliminated any sample that could not be scored on 20% of the SNPs attempted. We also removed data for any centre/SNP combination for which the call rate was less than 90%. In any instances where the call rate was 90–95%, the clustering of genotype calls was re-evaluated by an independent observer to determine whether the clustering was sufficiently clear for inclusion. We also eliminated all the data for a given SNP/centre where the reproducibility in duplicate samples was <97%, or where there was marked deviation from Hardy–Weinberg equilibrium in the controls ($P < 0.00001$).

Fine-scale mapping of *FGFR2*. Initial tagging of the associated region was done by identifying all SNPs with an m.a.f. > 5% in the HapMap CEPH/CEU set (Utah residents with ancestry from northern and western Europe). We then selected 7 SNPs (in addition to rs2981582) that tagged these variants with a

pairwise $r^2 > 0.8$, using the program Tagger (http://www.broad.mit.edu/mpg/tagger/)[37]. To identify additional common variants within the 32.5 kb region of linkage around the associated SNP, we resequenced 45 lymphocyte DNA samples from a subset of European subjects also genotyped by HapMap and other publicly available data sets. Seventy overlapping PCR amplicons were designed from positions 123317613 to 123348192 of chromosome 10 (average amplicon size 650 bp, 160 bp overlap). M13-tagged PCR products were bidirectionally sequenced using Big Dye 3.0 (Applied Biosystems) and processed using automated trace analysis through the Cancer Genome Workbench (cgwb.nci.nih.gov). Eighty-six per cent of the nucleotides across the region could be scored for polymorphisms in at least 80% of subjects. This set gave a >97% probability of detecting a variant with an m.a.f. > 5%. One hundred and seventeen variants were identified, including 27 present in dbSNP but without individual genotype information in European subjects, and an additional 46 not in dbSNP. Individual genotype information was then compared and merged with publicly available genotypes from Caucasian subjects (HapMap release 21 for 60 CEU parents, 22 European subjects from the Environmental Genome Project (EGP) resequencing effort (http://egp.gs.washington.edu/data/fgfr2/), and 24 European subjects from Perlegen (retrieved through http://gvs.gs.washington.edu/GVS)). There were 2 discrepancies among 389 genotype calls among subjects in common between our resequencing effort and EGP or Perlegen data, and 10 out of 926 compared to HapMap genotypes.

On the basis of these data, we identified 28 SNPs correlated with rs2981582 with $r^2 > 0.6$. We then attempted to genotype these 28 SNPs, plus rs2981582, in a subset of 80 controls from SEARCH and 84 controls from the Seoul Breast Cancer Study. Twenty-two of the variants were genotyped using Taqman. Four further variants (rs34032268, rs2912778, rs2912781 and rs7895676), which were not amenable to Taqman, were genotyped by Pyrosequencing (Biotage; http://www.biotagebio.com/). Assays were designed using Pyrosequencing Assay Design Software 1.0. The remaining 2 SNPs (rs35393331 and rs33971856) could not be genotyped using either technology and were excluded from further analyses. We cannot therefore comment on their likelihood of being the causal variant. Using these data, we selected tagging sets of 11 SNPs for UK subjects and 14 SNPs for Korean subjects (including rs2981582), such that each of the remaining variants was correlated with a tagging SNP with $r^2 > 0.95$ in the UK study or $r^2 > 0.86$ in the Korean study. After genotyping the 11 tag SNPs in SEARCH, two of these SNPs (rs4752569 and rs35012336) showed strong evidence against being the causative variant and were not considered further. The remaining 12 tag SNPs from the Korean subset were then genotyped in the samples from the IARC-Thai Breast Cancer Study, the Breast Cancer Study in Taiwan and the Multi-Ethnic Cohort (MEC), by Taqman.

Statistical methods. The primary test used for each SNP was a Cochran-Armitage 1 d.f. score test for association between disease status and allele dose. In the combined analysis, we performed a stratified Cochran-Armitage test. Stage 1 was given a weight of 4 in this analysis (corresponding to a weight of 2 in the score statistic), to allow for the expected greater effect size given the inclusion of cases with a family history. In the stage 3 analyses, each study was treated as a separate stratum, except for the MEC, in which the European American and Japanese American subgroups were treated as separate strata. For all studies except the MEC, individuals from a minor ethnic group for that study were excluded. Per-allele and genotype-specific odds ratios, and confidence intervals, were estimated using logistic regression, adjusting for the same strata. The summary odds ratios in Fig. 2 are based on the data from the stage 3 studies only, to avoid the bias inherent in estimates from the stage 1 and 2 data for SNPs exhibiting an association (the so called 'winner's curse'). The effects of genotype on family history of breast cancer (first degree yes/no) and bilaterality were examined by treating these variables as outcomes in a stratified Cochran-Armitage test.

To assess the global significance of the SNPs in stage 3, we computed the sum of the χ^2 trend statistics (excluding the 6 SNPs reaching genome-wide significance, plus rs2107425 as it was in LD with rs3817198) over those SNPs (17 of 23) for which the estimated odds ratios in stage 3 were in the same direction as the combined stage 1/stage 2[38]. Under the null hypothesis of no association, the asymptotic distribution of this statistic is χ^2 with n degrees of freedom, where n has a binomial distribution with parameters 23 and 1/2. The significance of this statistic was then assessed by computing a weighted sum of the tails of the relevant χ^2 distributions.

For the fine-scale mapping of the *FGFR2* locus, we first derived haplotype frequencies using the haplo.stats package in S-plus[39], separately for the European and Asian populations, using data from the case-control studies on whom the tag SNPs were typed plus the 164 control individuals on whom all SNPs were typed. These were used to impute genotype probabilities for each identified SNP in each individual. We then used an EM algorithm to fit a logistic regression model assuming that each SNP in turn was the causal variant, allowing for uncertainty in the genotypes of untyped SNPs, and hence to determine the likelihood that each SNP was the causal variant.

Coverage of the stage 1 tagging set was estimated using HapMap phase II as a reference. We based estimates on 2,116,183 SNPs with an m.a.f. of >5% in the CEU population. Of the SNPs successfully genotyped in stage 1, 187,663 were also on HapMap. For those SNPs not on HapMap, we identified 'surrogate' SNPs that were in perfect LD based on genotyping of 24 Caucasians by Perlegen Sciences (269,203 SNPs)[18]. To estimate coverage, we determined the best pairwise r^2 for each HapMap SNP and each tag SNP or a surrogate SNP, using the HapMap CEU data. This coverage was summarized in terms of the distribution of r^2 by allele frequency in 10 categories.

To estimate the power to detect each of the associations found, we computed the non-centrality parameter for the test statistic at each stage, based on the per-allele relative risk, allele frequency and r^2. This was used to estimate the power for a given r^2, based on a simulated trivariate normal distribution for the score statistics after each stage to allow for the correlations in the test statistics. We assumed a cut-off of $P < 0.05$ for stage 1, $P < 0.00002$ for stage 2 and $P < 10^{-7}$ for stage 3 (the first is slightly conservative, as more SNPs than this were actually taken forward). The overall power was obtained by averaging the power estimates for each r^2 over the distribution of r^2 obtained from the HapMap data, applicable to a SNP of that frequency.

The expected number of significant associations after stage 2 (Table 1) was calculated using a bivariate normal distribution for the joint distribution of the (weighted) Cochran-Armitage score statistics after stage 1 and after both stages, using a correlation of 0.525 between the two statistics (reflecting the weighted sizes of the two studies). These calculations were based on the 205,586 SNPs reaching the required quality control in stage 1. Of these, 11,313 reached a $P < 0.05$, of which 7,405 (65.5%) were successfully genotyped to the required quality control in stage 2. Thus the expected number reaching a given significance level with good quality control was calculated from the total number expected to reach this level × 65.5%. We adjusted the variances of the test statistics, separately for stages 1 and 2, using the genomic control method[22]. The adjustment factor, λ, was estimated from the median of the smallest 90% of the test statistics for SNPs typed in that stage, divided by the predicted median for the smallest 90% of a sample of χ^2_1 distributions (that is, the 45% percentile of a χ^2_1 distribution, 0.375).

36. Dean, F. B. *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl Acad. Sci. USA* **99**, 5261–5266 (2002).
37. de Bakker, P. I. W. *et al.* Efficiency and power in genetic association studies. *Nature Genet.* **37**, 1217–1223 (2005).
38. Tyrer, J., Pharoah, P. D. P. & Easton, D. F. The admixture maximum likelihood test: A novel experiment-wise test of association between disease and multiple SNPs. *Genet. Epidemiol.* **30**, 636–643 (2006).
39. Schaid, D. J., Rowland, C. M., Tines, D. E., Jacobson, R. M. & Poland, G. A. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am. J. Hum. Genet.* **70**, 425–434 (2002).